98P7687

3

B3

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平9－152886

(43)公開日　平成９年(1997)６月10日

(51)Int.Cl.⁶	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ１０Ｌ　3/00	５３５		Ｇ１０Ｌ　3/00	５３５	
	５２１			５２１Ｆ	

審査請求　有　　請求項の数５　ＯＬ　（全 24 頁）

(21)出願番号　特願平7－312286

(22)出願日　平成７年(1995)11月30日

(71)出願人　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地

(72)発明者　マリ・オステンドルフ
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(72)発明者　ハラルド・シンガー
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

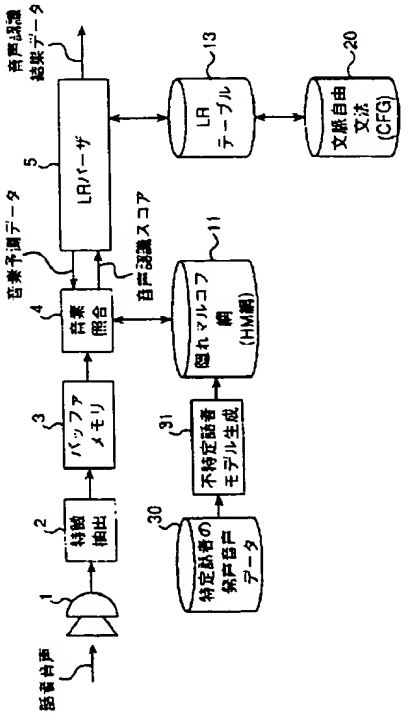
(74)代理人　弁理士　青山　葆　（外２名）

(54)【発明の名称】　不特定話者モデル生成装置及び音声認識装置

(57)【要約】

【課題】　従来例に比較して処理装置のメモリ容量の少なくてすみ、その計算時間を短縮することができる不特定話者モデル作成装置及び音声認識装置を提供する。

【解決手段】　複数の特定話者の発声音声データに基づいて、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布のＨＭＭを生成し、１つの状態をコンテキスト方向又は時間方向に分割したときに、最大の尤度の増加量を有する状態を検索する。次いで、最大の尤度の増加量を有する状態を、最大の尤度の増加量に対応するコンテキスト方向又は時間方向に分割した後、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布のＨＭＭを生成し、上記の処理を、単一ガウス分布のＨＭＭ内の状態を分割することができなくなるまで又は予め決められた分割数となるまで繰り返すことにより、話者独立型のＨＭＭを生成する。また、生成された話者独立型のＨＭＭを用いて音声認識する。



【特許請求の範囲】

【請求項１】　複数の特定話者の発声音声データに基づいて話者独立型の隠れマルコフモデルを生成するモデル生成手段を備えた不特定話者モデル生成装置において、

上記モデル生成手段は、複数の特定話者の発声音声データに基づいて、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布の隠れマルコフモデルを生成した後、上記単一ガウス分布の隠れマルコフモデルにおいて、１つの状態をコンテキスト方向又は時間方向に分割したときに、最大の尤度の増加量を有する状態を分割することを繰り返すことにより話者独立型の隠れマルコフモデルを生成することを特徴とする不特定話者モデル生成装置

【請求項２】　上記モデル生成手段は、

複数の特定話者の発声音声データに基づいて、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布の隠れマルコフモデルを生成する初期モデル生成手段と、

上記初期モデル生成手段によって生成された単一ガウス分布の隠れマルコフモデルにおいて、１つの状態をコンテキスト方向又は時間方向に分割したときに、最大の尤度の増加量を有する状態を検索する検索手段と、

上記検索手段によって検索された最大の尤度の増加量を有する状態を、最大の尤度の増加量に対応するコンテキスト方向又は時間方向に分割した後、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布の隠れマルコフモデルを生成する生成手段と、

上記生成手段の処理と上記検索手段の処理を、単一ガウス分布の隠れマルコフモデル内の状態を分割することができなくなるまで又は単一ガウス分布の隠れマルコフモデル内の状態数が予め決められた分割数となるまで繰り返すことにより、話者独立型の隠れマルコフモデルを生成する制御手段とを備えたことを特徴とする請求項１記載の不特定話者モデル生成装置.

【請求項３】　上記検索手段によって検索される状態は、直前の処理で上記生成手段によって分割された新しい２つの状態に限定されることを特徴とする請求項２記載の不特定話者モデル生成装置.

【請求項４】　上記検索手段によって検索される状態は、直前の処理で上記生成手段によって分割された新しい２つの状態と、上記新しい２つの状態から距離が１だけ離れた状態とに限定されることを特徴とする請求項２記載の不特定話者モデル生成装置.

【請求項５】　入力される発声音声文の音声信号に基づいて所定の隠れマルコフモデルを参照して音声認識する音声認識手段を備えた音声認識装置において、

上記音声認識手段は、請求項１乃至４のうちの１つに記載の不特定話者モデル生成装置によって生成された話者独立型の隠れマルコフモデルを参照して音声認識することを特徴とする音声認識装置.

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、複数の特定話者の発声音声データに基づいて不特定話者の隠れマルコフモデル（以下、ＨＭＭという）を生成する不特定話者モデル生成装置、及び、入力される発声音声文の音声信号に基づいて当該不特定話者のＨＭＭを参照して音声認識する音声認識装置に関する

【０００２】

【従来の技術】例えば、従来文献１　Ｊ．Ｔａｋａｍｉ　ｅｔ　ａｌ．，"Ａ　ｓｕｃｃｅｓｓｉｖｅ　ｓｔａｔｅ　ｓｐｌｉｔｔｉｎｇ　ａｌｇｏｒｉｔｈｍ　ｆｏｒ　ｅｆｆｉｃｉｅｎｔ　ａｌｌｏｐｈｏｎｅ　ｍｏｄｅｌｉｎｇ"，Ｐｒｏｃｅｅｄｉｎｇｓ　ｏｆ　ｔｈｅ　Ｉｎｔｅｒｎａｔｉｏｎａｌ　Ｃｏｎｆｅｒｅｎｃｅ　ｏｎ　Ａｃｏｕｓｔｉｃ　Ｓｐｅｅｃｈ　ａｎｄ　Ｓｉｇｎａｌ　Ｐｒｏｃｅｓｓｉｎｇ，ｖｏｌ．Ｉ，ｐｐ．５７３－５７６，１９９２年）において、従来例の逐次状態分割法（以下、ＳＳＳ法という）を用いたＨＭＭの生成方法が開示されており、当該ＳＳＳ法はＨＭＭ生成のために有効的な技術であり、最適なＨＭＭのトポロジーを自動的に学習するメカニズムを提供するものである　従来例のＳＳＳ法の背景にある基本的発想は、ＨＭＭにおける状態を分割するための最も大きな分散を有する状態を選択し、次いでその状態にとって最適な分割方向を採用すれば、ＨＭＭの状態ネットワーク（以下、ＨＭ網という）を増大させることができるというものである　この分割を繰り返し適用して得た結果がＨＭ網であり、特定の単語より小さいのサブワードの単位（例えば音素、モーラ）によるコンテキスト及び時間に対する分散性を効果的に表している　本出願人が行った幾つかの研究においてＳＳＳ法は成功裡に使用されており、また他のＨＭＭ設計を凌ぐ性能を示している（例えば、従来文献２　Ａ．Ｎａｇａｉ　ｅｔａｌ．，"Ｔｈｅ　ＳＳＳ法－ＬＲ　Ｃｏｎｔｉｎｕｏｕｓ　Ｓｐｅｅｃｈ　Ｒｅｃｏｇｎｉｔｉｏｎ　Ｓｙｓｔｅｍ：Ｉｎｔｅｇｒａｔｉｎｇ　ＳＳＳ－Ｄｅｒｉｖｅｄ　Ａｌｌｏｐｈｏｎｅ　Ｍｏｄｅｌｓ　ａｎｄ　ａ　Ｐｈｏｎｅｍｅ－Ｃｏｎｔｅｘｔ－Ｄｅｐｅｎｄｅｎｔ　ＬＲ　Ｐａｒｓｅｒ"，Ｐｒｏｃｅｅｄｉｎｇｓ　ｏｆ　Ｉｎｔｅｒｎａｔｉｏｎａｌ　Ｃｏｎｆｅｒｅｎｃｅ　ｏｎＳｐｏｋｅｎ　Ｌａｎｇｕａｇｅ　Ｐｒｏｃｅｓｓｉｎｇ，ｐｐ．１５１１－１５１４，１９９２年）、従来文献３　Ａ．Ｎａｇａｉ　ｅｔ　ａｌ．，"ＡＴＲＥＵＳ：Ａ　Ｃｏｍｐａｒａｔｉｖｅ　Ｓｔｕｄｙ　ｏｆ　Ｃｏｎｔｉｎｕｏｕｓ　Ｓｐｅｅｃｈ　Ｒｅｃｏｇｎｉｔｉｏｎ　Ｓｙｓｔｅｍｓ　ａｔ　ＡＴＲ"，Ｐｒｏｃｅｅｄｉｎｇｓ　ｏｆ　ｔｈｅ　Ｉｎｔｅｒｎａｔｉｏｎａｌ　Ｃｏｎｆｅｒｅｎｃｅ　ｏｎＡｃｏｕｓｔｉｃ　Ｓｐｅｅｃｈ　ａｎｄ　Ｓｉｇｎａｌ　Ｐｒｏｃｅｓｓｉｎｇ，ｖｏｌ．ＩＩ，ｐｐ．１３９－１４２，１９９３

年，及び、従来文献1 S. Sagayama et al., "ATREUS: a Speech Recognition Front-end for a Speech Translation System, Proceedings of European Conference on Speech Communication and Technology, pp. 1287-1290, 1993年 参照）

【0003】

【発明が解決しようとする課題】現在実施されているSSS法の不利な点は、これが良好に動作するのが話者依存型データのトポロジーの学習に限定されていることである 話者独立型の学習においては、SSS法によって選択される、最も分散の大きな状態は、調音結合の影響又は時間方向の影響よりもむしろ、話者の分散性をより反映するものと思われる SSS法を使用して話者独立型のモデルを構築するためには、従来文献5「T. Kosaka et al., "Tree-Structured Speaker Clustering for Speaker-Independent Continuous Speech Recognition", Proceedings of International Conference on Spoken Language Processing, pp. 1375-1378, 1994年」において開示されているように、まず話者依存型のトポロジーを設計し、次いでそれを話者独立型データについて再学習する。この解法は、プロの話者が注意深く読み上げる音声実験では良好に作用する（例えば、従来文献6「J. Takami et al., "Automatic Generation of Speaker-Common Hidden Markov Network by Adding the Speaker Splitting Domain to the Successive State Splitting Algorithm", Proceedings of Acoustic Society in Japan, pp. 155-156, 1992年」参照）が、管理の行き届かない状況においては、認識の点で限界があるものと思われる。特に自然に発話する発声音声においては、一人の話者にとって最適なトポロジーが、異なるアクセント、テンポ、スタイルの他の話者には適さない場合もある。

【0004】多くのHMMを用いた音声認識システムに使用されているもので、このSSS法に似た技術としては、時々決定木のコンテキスト方向のモデリングと呼ばれる、分割型分布クラスタリングがある。決定木の設計技術を利用した分布状態の分割型クラスタリングは、例えば、従来文献7「L. R. Bahl et al., "Decision trees for phonological rules in continuous speech", in Proceedings of the International Conferenceon Acoustic Speech and Signal Processing, pp. 185-188, 1991年」、及び、従来文献8「K.-F. Lee et al., "Allophone Clustering for Continuous Speech Recognition", Proceedings of the International Conference on Acoustic Speech and SignalProcessing, pp. 749-752, April 1990年」において全音素のモデルクラスタリング用として最初に提案され、後に連結された混合のための状態レベルのクラスタリング（例えば、従来文献9「X. Huanget al., "An overview of the SPHINX-II speech recognition system", Proceedings of ARPA Workshop on Human Language Technology, pp. 81-86, 1993年」参照。）、及び、単一ガウス分布へのクラスタリング（例えば、従来文献10「A. Kannan et al., "Maximum Likelihood Clustering of Gaussians for Speech Recognition", IEEE Transactions on Speech and Audio Processing, vol. 2, no. 3, pp. 453-455, 1994年」、従来文献11「S. J. Young et al., "Tree-based state tying for high accuracy acoustic modeling", in Proceedings of ARPA Workshop on Human Language Technology, pp. 307-312, 1994年」、及び、従来文献12「L. Bahl et al., "Context-dependent vector quantization for continuous speech recognition", Proceedings ofthe International Conference on Acoustic Speech and Signal Processing, vol. II, pp. 632-635, 1993年」参照。）に拡張されている これらの方法のアプローチは、ビタビ（Viterbi）のアルゴリズムスは前向き後向きアルゴリズムの何れかを使用して学習の観測データを、予め決定された幾つかのHMMのトポロジーが与えられた状態とを結合させ、次いで学習データの最大尤度に関する目的関数に

基づいてコンテキスト方向の分割のために決定木を成長させている．ＳＳＳ法とは違い、決定木のコンテキスト方向のモデリングは話者独立型ＨＭＭの学習において有効的に利用されている．決定木法とＳＳＳ法の両方法間の重大な相違点は、どの分布を分割するかの選択が、ＳＳＳ法では状態分布の分散の包括的な測定に基づいて行なわれるのに対して、決定木モデリングでは特定のコンテキスト方向の分割に基づいているという点である．

【０００５】上述のように、ＳＳＳ法のアルゴリズムは、図１０に示すＨＭ網を連続して成長させる繰り返しアルゴリズムである．ＨＭ網の各状態にはそれぞれ、下記の情報が割り当てられている．

（１）状態番号、（２）受理可能な異音カテゴリー（音素環境要因の直積空間として定義される）、（３）先行状態及び後続状態のリスト、（４）出力確率分布のパラメータ、及び、（５）自己遷移確率及び後続状態への遷移確率．このＨＭ網に対してある音素サンプルとその音素環境情報が与えられた場合、その音素環境を受理可能な状態を先行及び後続状態リストの制約内で始端から終端まで結ぶ１本の経路を一意に決定することができる．この経路に沿って各状態を連結したものは、図１１に示すようなＨＭＭと等価なモデルとなる．従って、経路選択後は通常のＨＭＭと同様に、出力尤度計算やモデルパラメータの推定のためのアルゴリズムをそのまま使用することができる．

【０００６】従来例のＳＳＳ法では、図１２に示すように、まず、２つの混合からどちらがより大きく発散しているかに基づいて分割すべき状態を選択し、次にコンテキスト方向（音素環境方向ともいう）及び時間方向への分割についてテストする．図４はそのトポロジーの変化を図示したものであり、選択された状態がコンテキスト方向及び時間方向に分割される様子を表している．

【０００７】従来例のＳＳＳ法を用いた話者モデル生成処理を示すフローチャートを図１３に示す．図１３において、まず、ステップＳ１で複数の特定話者の発声音声データに基づいて、公知のバーム・ウェルチの学習アルゴリズムを用いて混合数２の混合ガウス分布のＨＭ網を生成する．次いで、ステップＳ２で、最大の分散を有する分割すべき状態を検索して決定する．そして、ステップＳ３で決定された分割すべき状態に対してコンテキスト方向の分割テストと、２個の時間方向の分割テストを実行して、状態を分割するための最良の方向を決定する．さらに、ステップＳ４では、被影響状態を検索して決定し、Ｋ個の被影響状態に対してＫ個のコンテキスト方向のクラスタリングを実行することにより、各被影響状態に対して新しい混合ガウス分布の初期状態の混合分布パラメータを計算する．次いで、ステップＳ５では、被影響状態に対してバーム・ウェルチの学習アルゴリズムを用いて混合数２の混合ガウス分布のＨＭ網を生成する．そして、ステップＳ６で各状態が分割不可能であるか否かまたは予め決められた分割数（以下、所定の分割数という．）となったか否かが判断され、分割可能でありかつ所定の分割数に達していないならば、ステップＳ２に戻って上記の処理を繰り返す．一方、分割できないときまたは所定の分割数に達しているときは、ステップＳ７で得られたＨＭ網をメモリに格納する．

【０００８】「被影響状態」には、音素境界が固定されているものと仮定すれば、この分割によりパラメーターが変化する可能性のあるすべての状態が含まれる．音素境界の存在は、手動でラベル付けされたマークスはビタビの調整方法を用いて知ることができる．さらに詳細に言えば、被影響状態は、図４が示すように、ネットワークがダミー開始ノード及び終了ノードで切断された後に分割すべき現在の状態に連結される複数の状態のサブネットワーク内のすべての状態である．この定義によれば、より特定的な音素依存サブネットワークが展開されるまで、分割が行われる毎にほとんど全ての状態が影響を受けることになる．コンテキスト方向の分割の場合、コンテキスト依存上のミスマッチによって状態間の幾つかの新規パスが不可能になって取り除かれている場合がある点に注目する必要がある．図４の（ｂ）では、「×」がパスの「取り除き」を示している．

【０００９】従来例のＳＳＳ法のアルゴリズムにおける主な問題点は、分割すべき最適状態の選択が、実際の分割方向の選択に先立って行われることにある．各状態の出力分布は２つのガウス分布が混合されたものであり、「最適」状態とは、この２つの混合要素間の離散度が最大のものをいう．但し、こうした混合要素は包括的なものであり、可能な分割とは必ずしも対応していないため、可能な分割に制約があるとすれば、この基準による分割最適状態は事実上の最適な選択とはならない場合がある．例えば、話者独立型ＨＭＭの学習の場合、話者の多様性によって混合要素を十分に分離することは可能であるが、可能な分割が音声的なコンテキスト方向又は時間方向のものである場合は、新しい状態を加えることによって、この分散性をモデル化することはできない．その分割方法自体とは別に分割すべき状態を選択することにより、我々はまた非減少尤度の保証を失っている．但し、実際には、尤度の減少に出会うことはまれである．

【００１０】本発明の第１の目的は以上の問題点を解決し、多数の話者の膨大な学習用テキストデータを必要とせず、従来例に比較して処理装置のメモリ容量の少なくてすみ、その計算時間を短縮することができる不特定話者モデル作成装置を提供することにある．

【００１１】また、本発明の第２の目的は、上記第１の目的に加えて、生成された不特定話者モデルを参照して音声認識することができ、従来例に比較して音声認識率を改善することができる音声認識装置を提供することにある．

【００１２】

【課題を解決するための手段】本発明に係る請求項１記載の不特定話者モデル生成装置は、複数の特定話者の発声音声データに基づいて話者独立型の隠れマルコフモデルを生成するモデル生成手段を備えた不特定話者モデル生成装置において、上記モデル生成手段は、複数の特定話者の発声音声データに基づいて、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布の隠れマルコフモデルを生成した後、上記単一ガウス分布の隠れマルコフモデルにおいて、１つの状態をコンテキスト方向又は時間方向に分割したときに、最大の尤度の増加量を有する状態を分割することを繰り返すことにより話者独立型の隠れマルコフモデルを生成することを特徴とする。

【００１３】また、請求項２記載の不特定話者モデル生成装置は、請求項１記載の不特定話者モデル生成装置において、上記モデル生成手段は、複数の特定話者の発声音声データに基づいて、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布の隠れマルコフモデルを生成する初期モデル生成手段と、上記初期モデル生成手段によって生成された単一ガウス分布の隠れマルコフモデルにおいて、１つの状態をコンテキスト方向又は時間方向に分割したときに、最大の尤度の増加量を有する状態を検索する検索手段と、上記検索手段によって検索された最大の尤度の増加量を有する状態を、最大の尤度の増加量に対応するコンテキスト方向又は時間方向に分割した後、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布の隠れマルコフモデルを生成する生成手段と、上記生成手段の処理と上記検索手段の処理を、単一ガウス分布の隠れマルコフモデル内の状態を分割することができなくなるまで又は単一ガウス分布の隠れマルコフモデル内の状態数が予め決められた分割数となるまで繰り返すことにより、話者独立型の隠れマルコフモデルを生成する制御手段とを備えたことを特徴とする。

【００１４】さらに、請求項３記載の不特定話者モデル生成装置は、請求項２記載の不特定話者モデル生成装置において、上記検索手段によって検索される状態は、直前の処理で上記生成手段によって分割された新しい２つの状態に限定されることを特徴とする。

【００１５】さらに、請求項４記載の不特定話者モデル生成装置は、請求項２記載の不特定話者モデル生成装置において、上記検索手段によって検索される状態は、直前の処理で上記生成手段によって分割された新しい２つの状態と、上記新しい２つの状態から距離が１だけ離れた状態とに限定されることを特徴とする。

【００１６】またさらに、請求項５記載の音声認識装置は、入力される発声音声文の音声信号に基づいて所定の隠れマルコフモデルを参照して音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項１乃至４のうちの１つに記載の不特定話者モデル生成装置によって生成された話者独立型の隠れマルコフモデルを参照して音声認識することを特徴とする。

【００１７】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。

＜１．本実施形態の特徴＞図１は、本発明に係る一実施形態である不特定話者連続音声認識装置のブロック図である。本実施形態の音声認識装置は、特に、特定話者の発声音声データメモリ３０に格納された複数Ｎ人の特定話者の発声音声データに基づいて、従来例のＳＳＳ法を改良した話者独立型ＳＳＳ法（以下、ＳＩ－ＳＳＳ法という。）を用いて、不特定話者の話者独立型ＨＭ網１１を生成してそのメモリに格納する不特定話者モデル生成部３１を備え、ＨＭ網１１を参照して音声認識を行うことを特徴とする。この音声認識装置は、マイクロホン１と、特徴抽出部２と、バッファメモリ３と、音素照合部４と、文脈自由文法データベース２０内の所定の文脈自由文法に基づいて生成されたＬＲテーブル１３を参照して音声認識処理を実行する音素コンテキスト依存型ＬＲパーザ（以下、ＬＲパーザという。）５とを備える。

【００１８】＜２．ＳＩ－ＳＳＳ法の不特定話者モデル生成処理＞図２は、不特定話者モデル生成部３１によって実行される不特定話者モデル生成処理を示すフローチャートである。ここで我々は、「話者独立型ＨＭ網のトポロジー学習問題」に対して従来例のＳＳＳ法とは異なる解決方法を提案する。すなわち、単に状態にとって最適な分割法を求める段階、及び分割に最適な状態の摘出段階とを再配列する方法である。ＳＳＳ法と区別するため、ここではＳＩ－ＳＳＳ法と呼称するこの新アルゴリズムについて図２を参照して説明する。

【００１９】図２において、ステップＳ１１では、複数の特定話者の発声音声データ（具体的には、発声音声の特徴パラメータのデータである。）３０に基づいてそれぞれ後述する所定の音声の特徴パラメータを抽出した後音素を切り出して、従来の方法で複数の特定話者用単一ガウス分布のＨＭ網を生成する。そして、生成したＨＭ網に基づいて、公知のバーム・ウェルチの学習アルゴリズムを用いて学習を行って単一ガウス分布のＨＭ網を生成する。次いで、ステップＳ１２では、ＨＭ網内のすべての状態に対して分割可能な状態の分割情報を得る。この処理は、ステップＳ１５と同様に実行される。すなわち、詳細後述する最尤分割設定処理を用いてすべての状態に対して将来の分割の中で最良の分割方向及び音素（又は音素ラベル）を検索して決定し、これらを分割情報としてメモリに記憶する。すなわち、分割情報とは、以下の通りである。

（１）分割したときの期待される尤度の増加量、（２）分割は、コンテキスト方向であるか、時間方向であるか、並びに、（３）コンテキスト方向の前の音素、当該音素、後の音素

【００２０】次いで、ステップＳ１３において、分割情報に基づいて最大の尤度の増加量を有する分割すべき状

態を検索し、検索した状態を分割情報に従って分割する。すなわち、最大の尤度を有する分割すべき状態を最良の方向（すなわち、コンテキスト方向か、時間方向）で分割する。さらに、ステップＳ１４では、分割したときの被影響状態を検索して決定し、これらの被影響状態に対して公知のバーム・ウェルチの学習アルゴリズムを用いて学習を行って単一ガウス分布のＨＭ網を生成する。そして、ステップＳ１５で、詳細後述する最尤分割設定処理を用いて、ステップＳ１３で分割された２つの状態及び被影響状態に対して将来の分割の中で最良の分割方向及び音素（又は音素ラベル）を検索して決定し、これらを分割情報としてメモリに記憶する。ここで、Ｋ個の被影響状態に対して（Ｋ－１）個のコンテキスト方向の分割テストと１個の時間方向の分割テストが実行される。ステップＳ１６では、単一ガウス分布のＨＭ網内の状態が分割不可能であるか、又は単一ガウス分布のＨＭ網内の状態数が予め決められた分割数（以下、所定の分割数という。）となったか否かが判断され、分割可能でありかつ所定の分割数に達していないときはステップＳ１３に戻って上記の処理を繰り返す。一方、ステップＳ１６で分割が不可能であるとき、又は所定の分割数に達しているときは、ステップＳ１７で得られたＨＭ網１１をメモリに格納する。

【００２１】新しい状態のための初期状態の混合パラメータを計算するＳＳＳ法の処理の図１２のステップＳ４は、音素コンテキストの最適な分割法を見つけるステップに非常に良く似ている。初期化には、異なるコンテキスト用のサンプル平均値に対して実施するＶＱ（ベクトル量子化）学習手順が含まれている。これは、詳細後述する分割アルゴリズムに類似している。この段階を少し改良し、後のテストのために最適な分割からの利得を保存することによって、図１２のステップＳ３を効率良く省略し、同時により正確な走査を実現することができる。本実施形態のＳＩ－ＳＳＳ法のアルゴリズムのさらなる優位点は、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布に対して学習を実行する点である。これは混合ガウス分布の場合より遥かに早い速度で実行される。

【００２２】バーム・ウェルチの学習アルゴリズムによる学習がＳＳＳ法アルゴリズムよりＳＩ－ＳＳＳ法アルゴリズムにおいて遥かに高速であるにも関わらず、この２つの方法による計算コストは、同一規模になると予想される。全ての被影響状態が更新された場合、両アルゴリズムにおけるコンテキスト方向の分割テストの回数は本質的に同数である。すなわち、被影響状態数をＫと仮定した場合、ＳＳＳ法は（Ｋ＋１）回である一方、ＳＩ－ＳＳＳ法はＫ回である。ＳＩ－ＳＳＳ法のコンテキスト方向の分割テストは、従来例のＳＳＳ法の混合初期化段階より幾分高価であるが、これは最短距離よりクラスタリングの最尤規準の方を使用しているからである。但し、その差は僅かなものであり、また、この段階はＳＳＳ法の全体的な計算量からすると比較的小さい部分でしかない。また、本実施形態のＳＩ－ＳＳＳ法の時間方向の分割も、詳細後述されるように、被分割状態から結果的に生じる２つのガウス分布に対してバーム・ウェルチの学習アルゴリズムを用いた学習を必要とすることから、ＳＳＳ法の時間方向の分割において用いられる単一の前向きアルゴリズムパスに比較するとやはり経費が掛かるはずである。さらに、２回のＳＳＳ法による時間方向の分割テストに比べると、Ｋ回のＳＩ－ＳＳＳ法による時間方向の分割テストの方に可能性があるはずである。但し、時間方向の分割コストは、前向きアルゴリズムによるデータ処理量が小さく（単一状態に写像するのみ）、また、時間方向の分割は最大状態長の制約によって結果的に却下されることから、すべてのアルゴリズムのコストのほんの一部を占めるだけである。従って、ＳＩ－ＳＳＳ法の時間方向の分割の追加コストが問題となることはない。事実、詳細後述するように、本発明者による実験によれば、ＳＩ－ＳＳＳ法は、分離された２６２０単語の話者依存型学習においてはＳＳＳ法より早いことが示されている。

【００２３】たとえ、ＳＩ－ＳＳＳ法の電子計算機の処理時間がＳＳＳ法と同等か、又は僅かに早いだけであるとしても、ＨＭ網の生成コストを削減する利点は依然として存在する。被影響状態の部分集合（サブセット）に関するパラメータの再初期化（ＳＳＳ法用）、又は最適分割方法の再評価（ＳＩ－ＳＳＳ法用）を行なうだけで、ＳＳＳ法及びＳＩ－ＳＳＳ法両方のコスト削減が可能である。例えば、被影響状態に関しては以下の３つのレベルが指定可能である。

（Ａ）分割により生成される２つの新たな状態、（Ｂ）これら新たな２つの状態にすぐに近接する全ての状態、すなわち、分割された新たな２つの状態から距離１にある各状態、並びに、（Ｃ）その他のすべての被影響状態。言い換えれば、図２のステップＳ１５において対象となる状態を、上記セット（Ａ）のみにするか、上記セット（Ａ）及び（Ｂ）のみにするか、上記すべてのセット（Ａ）、（Ｂ）及び（Ｃ）としてもよい。

【００２４】従来例のＳＳＳ法においては、セット（Ｃ）に属する状態の混合パラメータを再設定することは必要のない場合がある。ＳＩ－ＳＳＳ法では分割による変更が最小限であることが予想されることから、セット（Ｃ）に当たる状態の分割パラメータを幾つか再推定することは理に適っている。電子計算機の使用を増やすためのＳＩ－ＳＳＳ法のオプションには以下のものが含まれる。

（１）同一の分割を保持し、分割の平均値と分散のみを更新して新しい利得を計算する。

（２）分割方向（例えば、左コンテキスト方向）を保持するが、早期の収束のために、前のコンテキストを用い

た分割アルゴリズムの初期化を行ってその方向におけるコンテキスト方向の最適分割を再評価する
（３）全般的な状態の再評価を行う
【００２５】注意を要する点は、２つの新しい状態については、可能な分割方法の全てに対して評価を行わなければならない、また全般的な再評価を行なわずに済ますことのできるのは、その他の被影響状態だけであるという２点である　被影響状態が完全に再評価されれば、改善された本実施形態のＳＩ－ＳＳＳ法のアルゴリズムはどの段階においても、同一のＨＭＭモデルから開始される従来例のＳＳＳ法のアルゴリズムに比べて学習データの尤度のより大きな増加を保証することになる．しかしながら、実際には完全な再評価を行なうことなく、かつかなりの低コストでも良い結果が達成される
【００２６】＜３．状態分割と条件付きＭＬ推定＞分割生成に使用される可能性のある最大の尤度に関しては、３つの一般的な目標関数が存在する　最も単純なものは、他の幾つかの研究（従来文献９、１０、及び１２）で実証されているように、学習データを幾つかの事前指定されたトポロジーにおける状態に位置調整し、次いでその結果である状態分布をクラスタ化して、データ及び所定の状態シーケンスの接続尤度を最大化する方法である　この方法は、多くの用途において成功裡に使用されてきたもので本質的にはビタビスタイルの学習であるが、バーム・ウェルチの学習アルゴリズムに関連した部分最適として知られている
【００２７】第２のオプションは、観測データの尤度を直接的に最大化することであるが、尤度の算定には発話境界のような固定点間における前向きアルゴリズムの駆動が必要となる．従って、直接的な尤度規準が有効であるのは、ＳＳＳ法における音素境界のような中間固定点が使用されている場合に限られる　分割尤度は、被分割状態が区分される固定された境界回数の範囲内で、全てのデータサンプル及び状態を前向きアルゴリズムを使用して算出する　分割結果の良好性の尺度には、それが真の尤度であり、隣接する状態に対する分割の影響を組み入れているという優位点がある．不利な点は、音素境界を必要とすることであり、特に、従来例のＳＳＳ法は、手動でラベル付けされた音素境界を付して使用されてきた　ビタビ法により調整された音素境界も、ほぼ同じように動作するものと思われるが、実験において検証されていない　但し、分割尤度規準の真の不利な点は単に、最も要求頻度の高いと思われるＳＩ－ＳＳＳ法において使用するには単に高価すぎることである
【００２８】この問題に対する我々の解答は、尤度より、標準的なバーム・ウェルチ学習アルゴリズムの背後にある期待値－最大値（Ｅｘｐｅｃｔａｔｉｏｎ　Ｍａｘｉｚａｔｉｏｎ：ＥＭ）のアルゴリズム（以下、ＥＭアルゴリズムという．）の概念（従来文献１３　Ａ．Ｐ．Ｄｅｍｐｓｔｅｒ　ｅｔ　ａｌ．，"Ｍａｘｉｍｕｍ Ｌｉｋｅｌｉｈｏｏｄ　ｆｒｏｍ　Ｉｎｃｏｍｐｌｅｔｅ Ｄａｔａ　ｖｉａ　ｔｈｅ　ＥＭ　Ａｌｇｏｒｉｔｈｍ"，Ｊｏｕｒｎａｌ　ｏｆ　ｔｈｅ　Ｒｏｙａｌ Ｓｔａｔｉｓｔｉｃａｌ　Ｓｏｃｉｅｔｙ，Ｖｏｌ．３７，Ｎｏ．１，ｐｐ．１－３８，１９７７年）を利用して、期待された対数尤度を最大化することである　ＥＭアルゴリズムの背後にある基本的な結果は、観測データy_1^T及び隠されたまたは非観測要素s_1^Tの期待された対数尤度の増加を意味している．この要素は、例えば、ＨＭＭの状態である　期待された対数尤度$Q(\theta \mid \theta^{(p)})$は次式で表わすことができ、ここで、$E_{\theta^{(p)}}\cdots$はパラメータ$\theta(p)$に関する対数尤度の期待値である
【００２９】
【数１】$Q(\theta \mid \theta^{(p)}) = E_{\theta^{(p)}} \log p(y_1^T, s_1^T \mid y_1^T, \theta)$
【００３０】ここで、最悪の場合でも、観測データ$L(\theta) = \log p(y_1^T \mid \theta)$の尤度には変化を与えない
【００３１】
【数２】$Q(\theta \mid \theta^{(p)}) \geq Q(\theta^{(p)} \mid \theta^{(p)}) = L(\theta) \geq L(\theta^{(p)})$
【００３２】ＨＭＭにおける条件付き独立の仮定により、期待された対数尤度は以下のように表すことができる
【００３３】
【数３】

$$
\begin{aligned}
&Q(\theta \mid \theta^{(p)}) \\
&= E[\log p(y_1^T, s_1^T \mid \theta) \mid y_1^T, \theta^{(p)}] \\
&= \sum_{s_1^T} p(s_1^T \mid y_1^T, \theta^{(p)}) \log p(y_1^T, s_1^T \mid \theta) \\
&= \sum_{s_1^T} p(s_1^T \mid y_1^T, \theta^{(p)}) \sum_t [\log p(y_t \mid s_t, \theta_{A(s)}) \\
&\quad + \log p(s_t \mid s_{t-1}, \theta_{B(s)})] \\
&= \sum_s \sum_t \gamma_t(s) \log p(y_t \mid s, \theta_{A(s)}) \\
&\quad + \sum_{s,s'} \sum_t \xi_t(s, s') \log p(s_t = s \mid s_{t-1} = s', \theta_{B(s)})
\end{aligned}
$$

【００３４】ここで、
【数４】$\gamma_t(s) = p(s_t = s \mid y_1^T, \theta^{(p)})$

【数5】$\xi_t(s, s') = P(s_t = s, s_{t-1} = s' \mid y_1^T, \theta')$

【００３５】数3の形式は、各状態s毎の分布パラメータ$\theta_{A(s)}$及び遷移確率$\theta_{B(s)}$の分離最大化を考慮したものである。これによって、我々は期待された尤度が増加するように単一の状態（又は、分割後の2つの状態）に対するパラメータを推定することができ、それによって観測データの尤度に減少のないことが保証される。

【００３６】特に、状態分割のHMM生成に際して、我々は$\gamma_t(s)$及び$\xi_t(s, s')$が全てのs＝s*に関し固定されているという条件にしたがってQ（$\theta \mid \theta'$）を最大化している。ここで、「s*」は、状態sの被分割状態（分割された状態）を表しており、「s'」は、状態sの1つ前の時間の状態を表わす。詳細後述するように、初期分割が適正に選択された場合、条件付き関数Q（$\theta \cdot \theta'$）は、s＝s*に依存する項が変化せず、従って、他の項に関わる尤度も減少できないために、その非減少が保証される。従って、L（θ）は、非減少として保証される。状態s*から状態s及び状態s_2への分割Sに対する期待された対数尤度の利得は以下のように求められる。

【００３７】

【数6】

$$
\begin{aligned}
&G(S)\\
&= \sum_{s=s_0, s_1} \sum_t \gamma_t(s) \log p(y_t \mid s, \theta_{A(s)})\\
&\quad - \sum_t \gamma_t(s*) \log p(y_t \mid s*, \theta_{A(s*)})\\
&\quad + \sum_{s=s_0, s_1} \sum_{s'=s_0, s_1} \sum_t \xi_t(s, s') \log a_{s,s'}\\
&\quad - \sum_t \xi_t(s*, s*) \log a_{s*,s*}
\end{aligned}
$$

【数7】

$$
\begin{aligned}
&G(S)\\
&= \sum_{s_0, s_1} N_1(s) \log p(y_t \mid s, \theta_{A(s)}) - N_2(s*) \log p(y_t \mid s*, \theta_{A(s*)})\\
&\quad + \sum_{s=s_0, s_1} \sum_{s'=s_0, s_1} N_2(s, s') \log a_{s,s'} - N_2(s*, s*) \log a_{s*,s*}
\end{aligned}
$$

【００３８】ここで、$a_{s,s'} = p(s_t = s \mid s_{t-1} = s', \theta_2 \theta)$であり、次式の通りである。

【００３９】

【数8】

$$N_1(s) = \sum_t \gamma_t(s)$$

【数9】

$$N_2(s, s') = \sum_t \xi_t(s, s')$$

【００４０】観測分布パラメータのみによる利得は、以下のように表すことができる。

【００４１】

【数10】

$$
\begin{aligned}
&Go(S)\\
&= \sum_{s=s_1, s_2} N_1(s) \log p(y_t \mid s, \theta_{A(s)}) - N_1(s*) \log p(y_t \mid s*, \theta_{A(s)})\\
&= 0.5 \left[\sum_{m=1}^{M} \left[N_1(s*) \log \sigma_m^2(s*) - N_2(s_0) \log \sigma_m^2(s_0) \right.\right.\\
&\quad \left. - N_1(s_1) \log \sigma_m^2(s_1) \right]\\
&\quad - N_1(s*) \log \left[1 + \{ N_1(s_0) N_1(s_1) \} / N_1(s*)^2 \right.\\
&\quad \left. \left. \cdot \sum_{m=1}^{M} (\mu_m(s_0) - \mu_m(s_1))^2 / \sigma_m^2(s*) \right] \right]
\end{aligned}
$$

【００４２】但し、分布は対角共分散によって記述され、下付き文字mはM次元ベクトル要素を表わすものと仮定する。この特別な利得の形式は、従来文献１０に記載された、結合された平均値と共分散の尤度規準を使用している（従来文献１４「T. W. Anderson. "An Introduction to Multivariate Statistical Analysis", J. Wiley & Sons, New Y

ork，１９８１年）の第１０節３項の結果に基づく。）コンテキスト方向の分割の場合、状態遷移確率は一定に保持され、数１０は期待された全体の利得をもたらす。一方、時間方向の分割では、期待された全体の利得は以下の式で求められる

【００４３】

【数１１】

$$\begin{aligned}&G_{\text{change in}}(S)\\&=G_0(S)-N_1(s*,s*)\log a_{s*,s*}+N_1(s,s_1)\log a_{s,s_1}\\&\quad+N_2(s,s_2)\log a_{s,s_2}-N_1(s,s_2)\log a_{s,s_2}\\&=G_0(S)-N_1(s*,s*)\log a_{s*,s*}+N_1(s,s)\log a_{s,s}\\&\quad+(N_2(s)-N_1(s,s_2))\log(1-a_{s,s_1})-N_1(s,s_2)\log a_{s,s_2}\end{aligned}$$

【００４４】数１０及び数１１によって得られる規準を使用すれば、分割方向の範囲内、及び範囲外の、また状態間での異なる分割候補を比較し、すべての学習セットの期待される尤度を最大化させる分割を選択することができる。ただ数１０及び数１１は、尤度自体の増加ではなく、期待される尤度の増加を示しているため、Ｓにわたる$G(S)$の最大化は、尤度が非減少であることのみを保証するものであり、必ずしも尤度を最大化させる分割を選択しているという点を保証するものではないということに注意すべきである。

【００４５】数１０及び数１１の尺度は、観測と状態の期待される結合尤度が増大するため、従来例のＳＳＳ法において観測尤度に基づく分割方向の選択に使用されたテストとは異なる形式となっている。さらに、これらの数１０及び数１１は、分割に最も適したノードを決定する際にもＳＳＳ法で使用される規準（従来文献１の式（１））とは異なる形式を採用しているが、この場合は、ＳＩ－ＳＳＳ法の規準の方が望ましい。従来例のＳＳＳ法の規準は、２つの包括的混合要素の間の距離の度合いであり、単一状態を有することに関連する尤度の利得ではない。まして、学習データの尤度の増加に関連づけられるようなものではない。

【００４６】分割によるＨＭＭ生成における状態尤度$\gamma_t(s)$及び$\xi_t(s,s')$の使用は、メモリ容量の増大化を意味している。メモリ容量を減少させるために、我々は従来例のＳＳＳ法で使用された技術を利用している。それは、音素境界（手動でラベル付けされた又はビタビ法により調整された）を使用して各回毎に非ゼロ確率を有する状態のセットを制約する（すなわち、$\gamma_t(s)$のサイズの減少させる）というものである。

【００４６】＜４．コンテキスト方向分割の効率的なサーチ＞従来例のＳＳＳ法は基本的には分割クラスタリングのアルゴリズムであるため、類似問題の処理における進歩（すなわち、決定木による設計（従来文献１５「L. Breiman et al., "Classification andRegression Trees", Wadsworth International Group, 1984年 参照。）から恩恵を享受することができる。決定木による設計における問題は、ＸからYを予測するための関数$Y=f(X)$を設計することである。Yが値$y\in R^n$をとる場合、当該関数は通常、回帰木と呼ばれる。また、$y\in\{1,\ldots,M\}$である場合には分類木と呼ばれる。決定木関数fは、Yを直接的に予測するよりむしろ、音声認識（従来文献１６「L. Bahl et al., "A tree-based statistical language model for natural language speech recognition", IEEE Transactions on Acoustic Speech, and Signal Processing, Vol. 37, No. 7, pp. 1001-1008, 1989年」参照。）で使用される木言語モデルの場合のように確率分布$Ph(y|X)=P(y|f(X))$を推定するために使用可能である。推定分布の解釈は、音声認識（例えば、従来文献１０、１１参照。）における分割分布クラスタリングの使用に対応しているため、決定木による設計方法をここで適用している。

【００４７】決定木による設計又は一般にいう分割クラスタリングにおける典型的なアプローチは、旺盛に成長するアルゴリズムであり、各段階で目的関数を最も進歩させる分割を行ないながら連続的に木を成長させている。このアルゴリズムは、可能性のある全ての木木の集、すべての可能な変数Ｘ（Ｘの要素）、及び変数Ｘ上で分割可能なすべての方法についてテストを行うことを要求している。変数Ｘ用の最適分割の選択が最も頻度の高いルーチンであることから、それが比較的高速であることが重要である。離散変数Ｘが値Ｊを持っている場合、テストすべきバイナリー分割（又は２分割）は約2^{J-1}個存在し、これはほとんど絶望的に高価である。従来文献１５のブレーマンほかは、Ｍ＝２である場合に関しては、素早い解答を与えている。後に、従来文献１７「P. A. Chou, "Optimal partitioning for classification andregression trees", IEEE Transactionson Pattern Analysis and Machine Intelligence, Vol. 13, No. 4, pp. 340-354, 1991年4月」において、ジョウは、さらに一般的な事例（Ｊ≧２及びＭ≧２）に対する高速分割設計用のアルゴリズムを提供している。ジョウのアルゴリズムは、単に多くの木設計の目標関数に対して局部

的な最適化を示すだけであるにも関わらず、M及びJに比例していることから、M＞2の場合には、こうしたパラメータの1つ又はその他方において指数関数的である。先に提案されたCART アルゴリズム（従来文献15）よりもさらに効率が良い。例えば音素モデルを使用するHM網の生成問題においては、Xは分割可能方向（例えば時間的、又は左、右又は中央音素コンテキスト方向）である絶対的な（無条件の）変数で構成されている。コンテキスト方向の何れに対しても、Xの値は音素ラベルであり、日本語ではN＝26音素である。従って、状態分割に関するHM網の生成問題は、決定木の無条件の質問設計に類似しており、可能性のある分割を効率的にサーチするためのアルゴリズムの恩恵を受けることができる。

【0048】我々は、以下でジョウの2分割アルゴリズム（従来文献17参照。）について再検討することから開始し、次いでこのアルゴリズムが最大のガウスログ尤度の目的関数にどのように適用されるかを示す。我々は「HMM生成への適用」を明確にするため、「ノード」ではなく「状態」、「木」ではなく「HM網」といった用語を使用して、標準的な決定木の専門用語（及び簡略表記法）を用いて説明する。標準的な決定木の設計との1つの相違点は、観測データを単一のノード又は状態に割り当てることではなく、異なった状態にある観測データの尤度を記述する確率分布が存在していることである。最初の議論を単純化するために、観測データはビタビ法による調整（学習）によって得ることができる唯一の状態に関連するものと仮定する。次いで、その結果をバーム・ウェルチの学習アルゴリズムに使用できるように拡張する方法を示す。

【0049】＜4．1　一般的な絶対的な分割生成アルゴリズム＞以下では、変数Xを使用して状態sを分割するためのジョウ（従来文献17参照。）による分割アルゴリズムについて説明する。仮に状態sに導く変数XがセットA_xを形成すると仮定しよう。我々はまず、観測データL（y，yh）を、HM網の（又は決定木）生成において最小化されるべき損失関数と定義することから開始する。変数yhはyの1つの表示であって、量子化、回帰又は直接分類の場合と同様に、Yと同じ空間における値を取ることが可能であり、もしくは、上記の木言語モデル及び分布クラスタリング例の場合と同様に、Yを表す確率分布とすることができる。HM網における状態sの不純性は、ある状態における期待される最小損失であり、以下の式で与えられる。

【0050】

【数12】$i(s)=E\cdot L(Y,\theta(s))\mid s$

【0051】ここで、E[f（Y）｜s]は、S＝sと仮定したときの条件付き期待値であり、$\theta(s)$は状態sのセントロイド（重心、質量重心）であり次式で表される。

【0052】

【数13】

$$\theta(s)=\underset{yh}{\mathrm{argmin}}\ E[L(Y,yh)\mid s]$$

【0053】発散d（s，yh）は、状態sを表すものとしてセントロイド$\theta(s)$の代わりにyhを使用するときと比較したときの、期待された損失における差である。

【0054】

【数14】

$$d(s,yh)=E[L(Y,yh)\mid s]-i(s)$$

【0055】状態sに対して分割を行うときにおいては、我々はまずi（s）を固定値とし、次式で表わされるi（s）を、J個のアレイに分割することによって達成される可能な最小不純度（これも固定値である。）として開始する。

【0056】

【数15】

$$i_1(s)=\sum_j p(x_j\mid s)\,i(x_j)$$

【0057】ここで、x_jは、コンテキスト方向のファクタXが取り得る可能な値である。状態s及びs_jへのバイナリー分割の不純度は以下の式で表される。

【0058】

【数16】

$$i_2(s)=\sum_{k=0,1} p(s_k\mid s)\,i(s_k)$$

【0059】いま、次式のようにおく

【0060】

【数17】

$$\begin{aligned}&i(s)-i_2(s)\\&=|i(s)-i_2(s)|+|i_2(s)-i_1(s)|\\&=\Delta_1+\Delta_2\end{aligned}$$

【0061】i（s）及び$i_2(s)$は定数であるため、それらの差異も固定値であり、旺盛に成長する分割設計においては公知であるように、Δ_1を最大にすることは、Δ_2を最小にすることと等価である。ジョウは、次式を示している。

【0062】

【数18】

$$\Delta_2=\sum_j p(x_j\mid s)\,d(x_j,\theta(\alpha(x_j)))$$

【0063】この式は、　Δ_2の最小化は量子化器の設

計問題として解釈することが可能である」ということを意味しており、その目標は、予想される発散を最小化するための「エンコーダ」$\alpha(x)$及び「デコーダ」又はセントロイド$\theta(s_k)$を設計することにある。ここで、エンコーダは、$k=0,1$に対してパーティション$A_k=\{x:\alpha(x)=s_k\}$として記述が可能である。この問題に対する局部的に最適な解は、K－平均値アルゴリズム、又はベクトル量子化（従来文献18「Y. Linde et al., "An algorithm for vector quantizer design", IEEE Transactions on Communication, vol. COM－28, pp. 84－95, 1980年1月」のためのリンデ・ブゾー・グレイ（Linde－Buzo－Gray）のアルゴリズムと類似する反復アルゴリズムを使用して求めることができる。すなわち、これは収束し、又は、平均的な損失の相対的な変化量が幾つかのしきい値より小さくなるまで、パラメータα及びθの再推定を反復することである。さらに明確に言えば、以下の2つのステップを実行することである。

【0064】（1）各xに対して、新しいエンコーダ$\alpha(x_i)^{(p+1)}$を見つける

【0065】

【数19】

$$\alpha(x_i)^{(p+1)}=\underset{i}{\mathrm{argmin}}\ d(x_i,\theta(s_i)^{(p)})$$

【0066】ここで、「argmin」は引数を最小にするiの値を示す関数であり、数19はk＝0，1のときに次式を与える。また同様に、「argmin」に代えて、「argmax」のときは、引数を最大にするiの値を示す関数である。

【0067】

【数20】$A_k^{(p+1)}=\{x_i:\alpha(x_i)^{(p+1)}=k\}$

【0068】（2）k＝0，1に対して新しいデコーダ$\theta(s_k)^{(p+1)}$を見つける

【0069】

【数21】

$$\theta(s_k)^{(p+1)}=\underset{\theta}{\mathrm{argmin}}\ E[L(Y,\theta)\mid s_k]=\underset{\theta}{\mathrm{argmin}}\sum_{x_i\in A_k^{(p+1)}}p(x_i\mid s_k)\,d(x_i,\theta)$$

【0070】無条件の予測の特別な場合に対して、ジョウの反復の分割アルゴリズムは、ナーダスほか（従来文献19「A. Nadas et al., "An iterative "flip－flop" approximation of the most informative split in the construction of decision trees", Proceedings of the International Conferenceon Acoustic Speech and Signal Processing, pp. 565－568, 1991年」によって提案された反復アルゴリズムと同様であるが、最小値Δ_2に対する最大値Δ_1の解釈によってこの2段階のメカニズムには差異がある。

【0071】ジョウ（従来文献17）は、例えば、回帰のための重み付け2乗エラー値（$y\in R^M$）及び分類のための重み付けギニ・インデックス（Gini index）及び対数尤度（$y^T=[00\ldots010\ldots0]$クラスMを指示するために第m列に1を有するM値のクラスインディケータ）等を含む種々の損失関数によってこのアルゴリズムが使用可能であることを示している。ここでは、ガウス分布によって分布が特徴づけられていると仮定して、特に最大の対数尤度の目的関数のためのアルゴリズムについて特に説明する。

【0072】＜4．2　最大ガウス対数尤度の使用＞ガウス分布のクラスタリングの問題に関して、$y\in R^M$は我々の音声認識のアプリケーションにおけるケプストラムベクトルに対応している。Xの各要素は可能な分割方向（例えば、左コンテキスト方向の音素ラベル）であり、Xは J 値の離散セットをとるXの要素である（例えば、26個の可能な日本語の音素である。）。我々は、平均値$\mu(s)$及び共分散行列$\Sigma(s)$によってパラメトリック・ガウス分布のHMMモデル$P(y\mid s)$を仮定する。従って、状態は$\theta(s)=(\mu(s),\Sigma(s))$によって表される。状態sに対応するXの可能な値の空間は、A_kによって特徴づけられていることを思い出させる。目標は、$A_p=A_q\cup A_r$である場合に、状態s_q及びs_rへの状態sの最適な分割を見つけることにある。

【0073】再度、一般的なアルゴリズムについて言及すると、予め決められた幾つかの観測データ$L(y,\theta)$に基づいて、$d(s,\theta)$及び最適デコーダを見つけるための式を決定する必要がある。目的が最大尤度（最尤）である場合は、観測データは、$L(y,\theta)=-\log P(y\mid\theta)$と表わすことができる。この目的関数のもとでは、数13は以下のようになる。

【0074】

【数22】

$$\theta(s)=\underset{\theta}{\mathrm{argmin}}\ E[L(Y,\theta)\mid s]=\underset{\theta}{\mathrm{argmax}}\ E[\log p(Y\mid\theta)\mid s]=\underset{\theta}{\mathrm{argmax}}\sum_{t:x_t\in A_k}\log p(y_t\mid\theta)$$

【0075】ここで、数22の第3式において、$t:x$

$x_t \in A_s$のΣは、x_tがA_sに属するときのtを変化して（log　p（x_t，θ））の和を計算するものである。我々は学習用データから学習しており、また真値P（y｜s）が未知であるために、ここでは経験的分布を使用している。このことは、標準的な最大尤度のパラメータ推定であり、これは、平均値μ（s）及び共分散Σ（s）を与えていることに留意する。従って、数21で表される発散は以下のようになる

【0076】

【数23】

$$
\begin{aligned}
&d(s,\theta)\\
&=E[L(Y,\theta)\mid s]-I(s)\\
&=-\sum_{t:x_t\in A_s}\log p(y_t\mid\theta)+\sum_{t:x_t\in A_s}\log p(y_t\mid\theta(s))\\
&=(1/2)\,[N_s\log|\Sigma|+\sum_{t:x_t\in A_s}(y_t-\mu)^t\Sigma^{-1}(y_t-\mu)\\
&\quad -N_s\log|\Sigma(s)|-\sum_{t:x_t\in A_s}(y_t-\mu(s))^t\Sigma(s)^{-1}(y_t-\mu(s))]
\end{aligned}
$$

【0077】ここで、N_sは、状態sを写像（又はマッピング）するときの観測回数であり、$\theta=(\mu,\Sigma)$である。上付きの記号「t」はベクトル転置行列を指し、|A|は行列Aの行列式を表す。単一のJ個の変数のための状態sにおけるバイナリー分割設定処理は次のように実行される。

【0078】図3に、図2のステップS15で用いる最尤分割設定処理を示す。図3において、まず、ステップS21で、2つの仮定状態の単一ガウス分布の分布パラメータに対して次式のように初期値を割り当てる。

【0079】

【数24】

$$\theta^{(0)}(s_0)=\theta(s)=(\mu(s),\Sigma(s))$$

【0080】この特別な選択によって、状態の1つが元の状態の分布パラメータを持っているため、ベクトル量子化器設計で使用される方法と同様に、尤度が増加することが保証される。次いで、ステップS22でパラメータpに1がセットされ、ステップS23で新しいパーティション$\{A_0^{(p)},A_1^{(p)}\}$（具体的には、分割された状態である。）を見つける。そして、ステップS23でパラメータjに1をセットし、ステップS25で次の数26が成立するか否かが判断される。

【0081】

【数26】

$$\sum_{t:x_t=x_j}\log p(y_t\mid\theta^{(p-1)}(s_0))\geqq\sum_{t:x_t=x_j}\log p(y_t\mid\theta^{(p-1)}(s_k))$$

【0082】ステップS25でYESであれば、ステップS26で各音素ラベルx_j，j＝1，．．．．Jに対して、パーティション$A_0^{(p)}$にx_jを割り当て、NOであれば、ステップS27でパーティション$A_1^{(p)}$にx_jを割り当てる。そして、ステップS28でパラメータjが個数Jであるか否かが判断され、Kでないときは、ステップS29でパラメータjを1だけインクリメントしてステップS25に戻り上記の処理を繰り返す。ステップS28でYESであるとき、ステップS30で、標準的な最大尤度パラメータ推定法によりセントロイド{$\theta^{(p)}(s_k)=(\mu^{(p)}(s_k),\Sigma^{(p)}(s_k))$：K＝0，1}を次式を用いて計算する。

【0083】

【数27】

$$\mu^{(p)}(s_k)=(1/N_k)\sum_{t:x_t\in A_k^{(p)}}y_t$$

【数28】

$$
\begin{aligned}
&\Sigma^{(p)}(s_k)\\
&=(1/N_k)\sum_{x_j\in A_k^{(p)}}\ \sum_{t:x_t=x_j}(y_t-\mu^{(p)}(s_k))(y_t-\mu^{(p)}(s_k))^t
\end{aligned}
$$

ここで、$N_k=\sum_{x_j\in A_k^{(p)}}N_j$

である。

【0084】また、N_jは｛t：$x_t=x_j$｝における要素の数であり、$N_0+N_1=N_s$である。次いで、ステップS31で第1の収束条件としてパーティションは変化しないか否かが判断され、変化するときはメインルーチンに戻るが、変化しないときはステップS32で第2の収束条件として、次の数29を満足するか否かが判断される。

【0085】

【数29】$(L^{(p)}-L^{(p-1)})\quad L^{(p-1)}<\eta$ここで、

【数30】$L^{(p)}=-N_s\log|\Sigma^{(p)}(s_s)|-N_s\log\ \Sigma^{(p)}(s_s)|$

【0086】ここで、ηは発見的に選択された収束のためのしきい値である。また、次式が満足することに注意する。

【0087】

【数31】L'=L''

【0088】ステップS32で、数29を満足するならば、メインルーチンに戻り、一方、数29を満足しないならば、ステップS33でパラメータnを1だけインクリメントしてステップS23に戻り上記の処理を繰り返す。

【0089】上記アルゴリズムの両段階に対しては、あらゆるデータポイントに関するログ確率を累積することよりも十分な統計量を使ってデータを示すことにより計算過程を保存することができる。特に、まず、対象となる変数Xが取る得る各xiにおいて対して、状態sに関連したデータyiについて記述する累積統計量を計算する。ここで、Niは$X=x_i$を有する状態sにおけるサンプル（フレーム）数を表すものとする。そして、次式のように、1次及び2次統計量を定義する。

【0090】

$$2N_i \log|\Sigma(s_0)| + \sum_{t:x_t=x_i} (y_t-\mu(s_0))'\Sigma(s_0)^{-1}(y_t-\mu(s_0))$$
$$\leqq 2N_j \log|\Sigma(s_1)| + \sum_{t:x_t=x_j} (y_t-\mu(s_1))'\Sigma(s_1)^{-1}(y_t-\mu(s_1))$$

【0093】記号法を簡単化するため、反復回数を示す上付きの記号（p）を省略する。総和の項は、数32及び数33により与えられる統計量を使用するため、以下のように簡単化することができる。

【0094】

【数35】

$$\sum_{t:x_t=x_j} (y_t-\mu(s_0))'\Sigma(s_0)^{-1}(y_t-\mu(s_0))$$
$$= \sum_{t:x_t=x_j} tr[(y_t-\mu(s_0))(y_t-\mu(s_0))'\Sigma(s_0)^{-1}]$$
$$= tr\left[\sum_{t:x_t=x_j} (y_t-\mu(s_0))(y_t-\mu(s_0))'\Sigma(s_0)^{-1}\right]$$
$$= tr[\Sigma(y_t y_t' - y_t\mu(s_0)' - \mu(s_0)y_t' + \mu(s_0)\mu(s_0)')\Sigma(s_0)^{-1}]$$
$$= tr[(S_j^2 - S_j^1\mu(s_0)' - \mu(s_0)(S_j^1)' + N_j\mu(s_0)\mu(s_0)')\Sigma(s_0)^{-1}]$$

【0095】ここで、恒等式$z^TAz=tr(zz^TA)$、及び追跡関数tr(・)が1次演算子であるという事実を使用した。これらの結果を数34に組み合えると、以下のようなテストの式が得られる。

【0096】

【数36】$2N_i\log|\Sigma(s_i)| + tr[(S_i^2 - 2S_i^1\mu(s_i)' + N_i\mu(s_i)\mu(s_i)')\Sigma(s_i)^{-1}] \leqq 2N_j\log|\Sigma(s_j)| + tr[(S_j^2 - 2S_j^1\mu(s_j)' + N_j\mu(s_j)\mu(s_j)')\Sigma(s_j)^{-1}]$

【0097】十分な統計量を使用するパラメータ再推定方程式は以下の通りである

【0098】

【数37】

$$\mu(s_k) = (1/N_k)\sum_{x_j\in A_k}\sum_{t:x_t=x_j} y_t = (1/N_k)\sum_{x_j\in A_k} S_j^1(s)$$

【数38】

$$\Sigma(s_k) = (1/N_k)\sum_{x_j\in A_k}\sum_{t:x_t=x_j} (y_t-\mu(s_k))(y_t-\mu(s_k))'$$
$$= (1/N_k)\sum_{x_j\in A_k}\sum_{t:x_t=x_j} (y_t y_t' - y_t\mu(s_k)' - \mu(s_k)y_t' + \mu(s_k)\mu(s_k)')$$
$$= (1/N_k)\sum_{x_j\in A_k} (S_j^2 - S_j^1\mu(s_k)' - \mu(s_k)(S_j^1)' + N_j\mu(s_k)\mu(s_k)')$$

【0099】尤度テスト、及びパラメータ再推定方程式の両方は、もし対角共分散であると仮定すれば簡単化さ

れる．クラスタ尤度テストを簡単化するために、以下の式が成立することに留意する必要がある

【0100】

【数39】

$$\log|\Sigma| = \sum_{m=1}^{M} \log\sigma_m^2$$

【数40】

$$\tr(\Sigma_A \Sigma_B^{-1}) = \sum_{m=1}^{M} \sigma_{A,m}^2 / \sigma_{B,m}^2$$

【0101】依って、新しい再分割テストは以下のようになる

【0102】

【数41】

$$C_0 + \sum_m \{S_{1,m}^2(s) - 2S_{1,m}^1(s)\mu_m(s_0) + N_1\mu_m(s_0)^2\} / \sigma_m^2(s_0)$$
$$\leqq C_1 + \sum_m \{S_{1,m}^2(s) - 2S_{1,m}^1(s)\mu_m(s_1) + N_1\mu_0(s_1)^2\} / \sigma_m^2(s_1)$$

ここで、

【数42】

$$C_k = N_1 \log\left(\prod_m \sigma_m^2(s_k)\right)$$

【0103】また、共分散行列が対角行列であると仮定した場合、数38は次式のように簡単化される．

【0104】

【数43】

$$\sigma_m^2(s_k) = (1/N_k) \sum_{x_t \in A_k} (S_{1,m}^2 - 2S_{1,m}^1 \mu_m(s_k) + N_1 \mu_m(s_k)^2)$$

【数44】

$$\sigma_m^2(s_k) = \mu_m(s_k)^2 + (1/N_k) \sum_{x_t \in A_k} (S_{1,m}^2 - 2S_{1,m}^1 \mu_m(s_k))$$

ここで、m＝1，．．．．Mである

【0105】このアルゴリズムを、ビタビ・アルゴリズムではなくバーム・ウェルチの学習アルゴリズムを介して、観測データが状態に蓋然的に（見込みに基づいて）関連している場合にまで拡張するために、更新されている状態に存在する尤度によって、単に数32及び数33の和の内側の各項を単に重み付けする．特に、$\gamma_t(s)$を時間tのときの状態をsであるときの確率に対応させる．このとき、新しい十分な統計量は次式で表される

【0106】

【数45】

$$S_1^t(s) = \sum_{t: x_t = x_l} \gamma_t(s)\, y_t$$

【数46】

$$S_1^2(s) = \sum_{t: x_t = x_l} \gamma_t(s)\, y_t y_t^t$$

【数47】

$$N_1(s) = \sum_{t: x_t = x_l} \gamma_t(s)$$

【0107】$\gamma t(s)$項は、前向き及び後向きパスの両方を使用して計算するものとする．原理上、この情報は、ＳＳＳ法及びＳＩ－ＳＳＳ法におけるバーム・ウェルチの反復から利用可能であるが、ＳＳＳ法にはこの情報を全て格納するためのデータ構造がなく、ＳＩ－ＳＳＳ法の場合はそれを付加する必要がある．ＳＳＳ法では、分割に最適なノード、及びノードが決定されている場合の最適な分割方法の何れかを求めるに当たって$\gamma t(s)$の項を必要としない．これは、分割に最適なノードは包括的な混合ガウス分布から選択され、最良の分割は前向きアルゴリズムを使用するためである．

【0108】＜5．時間分割の制約つきＨＭ網の生成＞ＨＭ網の生成における我々の目標は、前述されたサーチ方法を用いて、各ステップ毎に学習用データの尤度を最大にまで増加させることである．我々は、上述のように、制約つきＥＭ（期待値－最大値）アルゴリズムの方法を肯定する議論を行ったが、これは、ＨＭ網に予想される尤度の増加量が、全体でも被分割状態と２つの新しい状態の期待される尤度の差に過ぎないというものである．並行する２つの状態の尤度は和をとれば元の状態の尤度になることから、コンテキスト方向の分割において前後の方向数を制限することは、すなおな方法である．しかしながら、シーケンス内の２つの状態の尤度は、単純な和によって与えられない．

【0109】本実施形態のＳＳＳ法において、時間方向分割は、ＨＭＭの前向きアルゴリズムの使用、及び分割された状態以外の状態の状態尤度（$\gamma t(s)$：時間tにおける状態sの尤度）の変更を含む．この場合、ネットワークの大部分は、ＨＭ網の全体としての尤度における変化を確立するように評価する必要がある．できる限

り大きなサブネットワークを評価するために必要な追加コストに加えて、他の状態の尤度が、時間方向の分割に対して変化するが、コンテキスト方向で変化しないという問題は、時間の方向で分割することを選択する方向に向かうバイアスとなるであろう。

【0110】制約つきEM基準は、時間方向の分割の設計におけるこうした問題点を処理するものであり、被分割状態以外の状態の尤度は分割設計のパラメータ推定段階において変化しないという制約がある。より明確化するために、図5のように状態s＊を分割された状態とし、状態q_0及びq_1を時間方向の分割によって得られる2つの状態とする。これらの関係をより明確にするため、仮説の新しい状態をqとし、分割された状態候補をs＊と表記する。新しい状態を記述するために推定しなければならないパラメータΘは、$\Theta=\{\mu(q_0), \sigma(q_0), \nu(q_0), \mu(q_1), \sigma(q_1), \nu(q_1)\}$である。ここで、$\mu(q)$は状態qの平均値ベクトルであり、$\sigma(q)$は分散量ベクトルであり、$\nu(q)$は状態qから状態qへの復帰確率、すなわち、セルフループ状の遷移確率を意味する。HM網におけるこうしたパラメータのみが変動し、他は変動しないという点を保証するためには、次のような制約が必要である。

【0111】

【数48】
$$\gamma_t(s*)=\gamma_t(q_0)+\gamma_t(q_1)$$
$$\xi_t(q,q')=p(q_t=q, q_{t-1}=q' \mid Y)$$
$$=p(q_t=q, q_{t-1}=q', s_t=s*, s_{t-1}=s* \mid Y)$$
$$=\xi h_t(q,q')\,\xi_t(s*,s*)$$

【0114】項$\gamma t b(q)$及び$\xi t b(q,q')$は、$\gamma t(s*)\neq 0$であるデータのみを使用し、かつ$\gamma t b(q_0)+\gamma t b(q_1)=1$となるような状態$q_0$及び$q_1$に対してのみの非ゼロ状態尤度を有する前向き－後ろ向き標準アルゴリズムを用いて計算することができる。従って、前向き－後ろ向きアルゴリズムを制約することは、単に前向きパス及び後ろ向きパスを適正に初期化し、もしくは、図6のハッチング部分として表されているようなすべてのデータ構造のサブセットを通過させるということに過ぎない。図6は、$\gamma h_t(q)$及び$\xi h_t(q,q')$を時間方向の分割に対して計算するときに用いるデータと状態とを示し、ここで、図6において、0は、不可能な状態を示している。一旦、項$\gamma t b(q)$及び$\xi t b(q,q')$が計算されると、次式に従ってパラメータΘを計算する。

【0115】

【数56】
$$\mu_s(q)=\{\Sigma_t \gamma h_t(q)\,\gamma_t(s*)\,y_{t,s}\}/\{\Sigma_t \gamma h_t(q)\,\gamma_t(s*)\}$$

【数57】
$$\sigma_s(q)=\{\Sigma_t \gamma h_t(q)\gamma_t(s*)y_{t,s}^2\}/\{\Sigma_t \gamma h_t(q)\gamma_t(s*)-\mu m(q)^2\}$$

【数58】
$$\nu(q)=\{\Sigma_t \xi h_t(q,q)\xi_t(s*,s*)\}/\{\Sigma_{q'}\Sigma_t \xi h_t(q',q)\xi_t(s*,s*)\}$$

【0116】ここで注意すべきことは、項$\gamma t b(q)$及び$\xi t b(q,q')$の計算に使用される前向き・後ろ向きアルゴリズムは、2つの新しい状態に写像する（マッピングされる）観測データの尤度を見つけるため

【数19】$\xi_t(s*,s*)=\xi_t(q_0,q_0)+\xi_t(q_0,q_1)+\xi_t(q_1,q_0)+\xi_t(q_1,q_1)$

ここで、

【数50】$\gamma_t(i)=p(s_t=i \mid Y)$

【数51】
$$\xi_t(i,j)=p(s_t=i, s_{t-1}=j \mid Y)$$

【0112】ここで、上記の式はHMMの再推定に必要な一般的な項であり、Yはすべての学習セットを表している。これらの制約条件は、数52及び数53を定義しかつ条件つき確率及び冗長性$s t=s*$の定義を用いることにより、容易に満足させることができ、数54及び数55を得ることができる。

【0113】

【数52】
$$\gamma h_t(q)=p(q_t=q \mid s_t=s*, Y)$$

【数53】$\xi h_t(q,q')=p(q_t=q, q_{t-1}=q' \mid s_t=s*, s_{t-1}=s*, Y)$

【数54】
$$\gamma_t(q)=p(q_t=q \mid Y)=p(q_t=q, s_t=s* \mid Y)=\gamma h_t(q)\,\gamma_t(s*)$$

【数55】

に用いることができない．従って，尤度における相対的な変化は，時間方向の分割の再学習のための停止基準として用いることができない．分割は後にバーム・ウェルチの学習アルゴリズムによって再学習されるため，単に固定された数の再学習の反復を実行することが妥当であり，ここでは4回の反復を使用した．

【0117】時間方向の分割によるHMM生成における問題は，コンテキスト方向の分割とは異なり，非減少尤度の保証がない点である．SI－SSS法による時間方向の分割再推定方法の場合は，それがEMアルゴリズムの制約つきバージョンであるために非減少尤度が保証されているが，1つの状態から2つの状態への分割においては，尤度が減少しないことを保証するように初期設定することができない．この処理において採用している適当な初期推定は，元の状態の観測分布を使用し，かつ，仮説の2つの状態における期待される持続時間の和が元の状態の期待される持続時間と同一になるように，遷移確率を選択するものである．実際には，まれではあるが時に尤度の減少は生じる．この場合，その状態の時間方向への分割は絶対に選択されない．SSS法の時間方向分割アルゴリズムも同様の問題を抱えている．これは，ガウス分布の固定されたセット用に最適の時間方向の分割を選択するからであるが，このガウス分布は時間方向の分割を特に想定して設計されたものではないため，こういった分割には必ずしも整合しない．しかしながら，ノードは混合要素間の距離に基づいて，実際の結果に関わらず分割されるため，従来例のSSS法のアルゴリズムは，悪い時間方向の分割を回避することができない．もちろん，SSS法のアルゴリズムは，分割処理において状態の再調整を実行することによって，SI－SSS法の時間分割より大きな中間利得を達成する可能性を有しているが，この差は小さいと考えられる．なぜなら，SI－SSS法が直後に続くバーム・ウェルチの再推定処理において状態の再調整を可能にしているためである．従って，結果的には，SI－SSS法による時間方向の分割テストの方が，SSS法の場合より改善されている．

【0118】≪6．本実施形態の不特定話者連続音声認識装置≫本実施形態においては，音声認識のための統計的音素モデルセットとしてHM網11を使用している．当該HM網11は効率的に表現された音素環境依存モデルである．1つのHM網は多数の音素環境依存モデルを包含する．HM網11はガウス分布を含む状態の結合で構成され，個々の音素環境依存モデル間で状態が共有される．このためパラメータ推定のためのデータ数が不足する場合も，頑健なモデルを生成することができる．このHM網11は，従来例のSSS法から改善されたSI－SSS法を用いて自動生成される．上記SI－SSS法ではHM網11のトポロジーの決定，異音クラスの決定，各々の状態におけるガウス分布のパラメータの推定を同時に行なう．本実施形態においては，HM網11のパラメータとして，ガウス分布で表現される出力確率及び遷移確率を有する．このため認識時には一般のHMMと同様に扱うことができる．さらに，上記HM網11を用いた，SSS－LR（left-to-right　rightmost型）不特定話者連続音声認識装置について説明する．この音声認識装置は，メモリに格納されたHM網11と呼ばれる音素環境依存型の効率のよいHMMの表現形式を用いている．

【0119】図1において，話者の発声音声はマイクロホン1に入力されて音声信号に変換された後，特徴抽出部2に入力される．特徴抽出部2は，入力された音声信号をA／D変換した後，例えばLPC分析を実行し，対数パワー，16次ケプストラム係数，Δ対数パワー及び16次Δケプストラム係数を含む34次元の特徴パラメータを抽出する．抽出された特徴パラメータの時系列はバッファメモリ3を介して音素照合部4に入力される．

【0120】音素照合部4に接続されるHM網11は，各状態をノードとする複数のネットワークとして表され，各状態はそれぞれ以下の情報を有する．

（a）状態番号
（b）受理可能なコンテキストクラス
（c）先行状態，及び後続状態のリスト
（d）出力確率密度分布のパラメータ
（e）自己遷移確率及び後続状態への遷移確率

【0121】音素照合部4は，音素コンテキスト依存型LRパーザ5からの音素照合要求に応じて音素照合処理を実行する．そして，不特定話者モデルを用いて音素照合区間内のデータに対する尤度が計算され，この尤度の値が音素照合スコアとしてLRパーザ5に返される．このときに用いられるモデルは，HMMと等価であるために，尤度の計算には通常のHMMで用いられている前向きパスアルゴリズムをそのまま使用する．

【0122】一方，文脈自由文法データベース20内の所定の文脈自由文法（CFG）を公知の通り自動的に変換してLRテーブル13を生成してそのメモリに格納される．LRパーザ5は，上記LRテーブル13を参照して，入力された音素予測データについて左から右方向に，後戻りなしに処理する．構文的にあいまいさがある場合は，スタックを分割してすべての候補の解析が平行して処理される．LRパーザ5は，LRテーブル13から次にくる音素を予測して音素予測データを音素照合部4に出力する．これに応答して，音素照合部4は，その音素に対応するHM網11内の情報を参照して照合し，その尤度を音声認識スコアとしてLRパーザ5に戻し，順次音素を連接していくことにより，連続音声の認識を行い，その音声認識結果データを出力する．上記連続音声の認識において，複数の音素が予測された場合は，これらすべての存在をチェックし，ビームサーチの方法により，部分的な音声認識の尤度の高い部分木を残すとい

う枝刈りを行って高速処理を実現する。

【0123】以上の実施形態において、特徴抽出部2と、音素照合部4と、LRパーザ5と、不特定話者モデル生成部31とは、例えばディジタル計算機によって構成される。また、特定話者の発声音声データ30とHMM網11と文脈自由文法データベース20とLRテーブル13とが格納される各メモリとバッファメモリ3とは、例えばハードディスクメモリで構成される。

【0124】

【実施例】

<7．実験>本実施例では、本実施形態のSI－SSS法の有効性を確立するための事前作業に相当する2つの実験について説明する。実験に使用したコーパスについて説明した後、読み上げ音声に対する話者依存型音声認識の実験結果について述べる。この実験は、SSS法が開発されたときに用いられた典型的な実験であり、極度に制御されたコーパスではSI－SSS法が解決しようとしているタイプの問題が現れないために、SI－SSS法にとっては最も難しいテストとなっている。次いで、我々の目標である話者独立型音声認識により近いタスクにおける利得を実証する予備段階として、多数話者タスクの実験を行った結果について説明する。

【0125】<7．1　典型的な実験>本発明の目的は、発声音声としてより良い話者独立型HMMを開発すること、話者独立型音声認識装置の動作において可能な限り最適な性能を獲得すること、及び話者適合化のより良い開始点を提供することである。本発明を発明したときに可能であった発声音声データ量がバーム・ウェルチの学習アルゴリズムを使用する学習には不十分であったため、我々は実際の読み上げ音声に対して初期話者モデルを学習させ、発声音声に対してベクトル場平滑化を用いてそのモデルを適合化することに重きを置いた。

【0126】これらの実験には、数種類のコーパス（言語資料）を使用した。話者独立型実験（話者1－2名）における初回のトポロジー学習と、話者依存型実験（話者6名）では、最も使用頻度の高い5240単語（Aセット）からなる分離された単語の読み上げによる日本語コーパスが使用されている。すべての読み上げ音声コーパスは、音素ラベルのシーケンスと、各音声セグメントに対する開始点と終了点とを用いて手書きで書き写されたものであり、このことにより、SSS法の学習を容易にしている。音声は、低ノイズ下でプロの話者により録音された。話者独立型実験では、Aセットで設計されたプロトタイプモデルを、話者独立型読み上げ音声データベース（Cセット：15名の話者が50音素を均等に配置した文章を異なる発話速度で3回発音している。）のサブセット上で再学習させた。Cセットデータの録音には、Aセットデータの場合と同じタイプのマイクロホンを使用した。ポーズ単位（休止単位）は手書きでマークが付けられたが、音素境界にはマークが付与されていない。Aセットと Cセットコーパスに関しては、従来文献20「A. Kurematsu et al., "ATR Japanese speech database as a tool of speech recognition and synthesis", SpeechCommunication, 9：357－363, 1990年」において説明されている。最後に、発声音声コーパスを使用して、音声認識の性能評価を行った（発声音声コーパスは、例えば、従来文献21「N. Uratani et al., "ATR Integrated Speech and Language Database", ATR Technical Report TR－IT－0056, 1991年」参照）。このコーパスは、従来文献22「H. Singer et al., "Development testsets and administrative tools for ATR's non-read speech databases (SLDB and SDB)", Technical Report TR－IT－0118, ATR, 1995年」において明記されているように、学習セットとテストセットとに分割した。すべての発声音声学習データは、日本語話者1名、英語話者1名、及び翻訳者（音声言語パート）2名を含む2言語の会話の集合からなる。データを収集したのは異なる3つの会社であるために、その品質（例えば、SNR（信号雑音比））にはかなりの差がある。使用するのは翻訳者でない話者による音声のみとし、メモリ上の制約から、少数の6秒以上のポーズ単位は学習セットから除外した。

【0127】話者独立型実験に使用する開発テストデータには、発声音声コーパスの音声言語パート及び音声パートの両方（すなわち、日本語話者2名間の単一言語会話）を使用した。このデータは、こうした発音には埋まったポーズが多いという点から、より「自然な」ものとなっている。このテストセットには、女性話者4名（会話形文体15、音素9711）と男性話者3名（会話形文体16、音素11231）が含まれている。

【0128】分析パラメータは、サンプリング周波数12000Hz、フレームシフト5ms、フレーム長20ms、プリエンファシス0．98、16次のLPC分析及び16次ケプストラム計算、16次Δケプストラム、パワー値、及びΔパワー値であった。Δケプストラム計算用の三角形の回帰窓の長さは、両面とも9フレーム分（すなわち、90msの両面窓）であった。認識実験は、日本語を音素対文法で表現した、音素配列上に制約のあるワンパス－ビタビアルゴリズムを使用して行った（従来文献23「H. Singer et al., "Speech recognition without grammar or vocabulary constraints", Proceedings

of International Conference on Spoken Language Processing, pp. 2207－2210, 1994年，参照）

【0129】<7．2　話者依存型HM網実験>SI－SSS法アルゴリズムが、少なくとも常にSSS法アルゴリズムと同等の性能を示すことを証明するため、我々は、初回の実験を話者依存型のモードで行った。200及び400状態数の単一ガウス分布HMMと混合数3の400状態ガウス分布HMMを、各話者毎にAセットの偶数番単語（2620語）について学習を行った。初回のトポロジーでは26の状態を使用して（図7が示すように、21音素のそれぞれの中心に各1状態と、全音素が共有する形の左右各1状態）、初回のHM網の学習時間を減少させ、また各音素が認識可能となることを保証した。これら複数のHMMは、1310語の奇数番単語についてテストされた。HM網のトポロジー生成後は、最大21の反復のバーム・ウェルチ反復を実行し、単一ガウス分布HMMの状態観測分布を推定した。単一ガウス分布HMMの場合、相対的な尤度利得に関するしきい値テストを用いて、通常10回未満の反復を必要とする。この実験結果は表1の通りである。

【0130】

【表1】

パーセントアキュラシー

話者	200状態1混合 SSS	200状態1混合 SI-SSS	400状態1混合 SSS	400状態1混合 SI-SSS	400状態3混合 SSS	400状態3混合 SI-SSS
MHT	93.9	92.8	95.4	94.5	96.1	96.0
MAU	93.6	93.2	95.2	95.2	96.4	96.7
MXM	91.7	91.9	93.6	93.9	95.3	95.1
FTK	91.5	91.1	92.9	94.0	94.7	95.0
FMS	89.7	91.3	91.9	93.2	94.2	94.6
FYM	90.7		92.4		92.9	
	93.6		95.1		95.5	
平均値	91.9	92.1	93.7	94.1	95.3	95.5

【0131】表1から明らかなように、平均的な場合（大部分がこの場合である）では、本実施形態のSI－SSS法の方が従来例のSSS法より僅かに良い結果を示していることが解る。唯一の例外が話者MHTであり、ほとんどのSSS法の開発作業に使用されたものである。特に、話者がプロでありまた録音が高品質であることから、この話者依存型データの状態分布の不要な分散はコンテキスト方向のものであり、その差は予想通り僅かなものである。

【0132】さらに、特に、状態数200のトポロジーの場合、本実施形態のSI－SSS法は、従来例のSSS法よりも多くの音素にわたって多数の異音を分配（又は分類）することがわかった。SI－SSS法は子音に対してSSS法より多くの異音を分配し、また、母音に関しては異音を幾分より均等に分配している。分配の相違点は、特に、「a」と「u」の対比において顕著であり、SSS法は「u」よりも「a」に関して格段に多くの異音を分配するが、SI－SSS法の場合、状態数400のHMMでは同様の数の異音を有するが、状態数200のHMMでは「u」の方に多くの異音を有している。

【0133】次いで、SSS法及びSI－SSS法のCPUの計算時間を測定した。計算時間は、SI－SSS法の方がSSS法より格段に短かった。これは、SI－SSS法の場合、時間方向の分割の方がコスト高であることから、特に可能性のあるすべての時間方向の分割が選択された後にこれが顕著である。シーケンス内の最大状態数が、最小の音素持続時間の制約条件である20msを有効に確立させるために、4に制限されているため、ここでは、時間方向分割数が限定されている。

【0134】図8は、話者FTKに関するCPUの計算時間の差を図示したものである。一方、本実施形態のSI－SSS法は、分割に使用するためバーム・ウェルチの状態尤度を格納しなくてはならず、SSS法より以上の記憶装置を必要とする。2620単語の学習セットでは、そのコストの差はおよそ、80MBに対して50MBである。話者10名であって、話者1名当たり1000語の話者独立型学習は、100MBの主記憶装置を使用し、パラメータファイルをディスクへスワップすることによって実行されるものと推定される。

【0135】<7．3　複数話者HM網実験>次に、図1の連続音声認識装置を用いて、複数話者の音声認識実験を行い、従来例のSSS法と本実施形態のSI　SSS法を比較した。6名の話者（MAU、MHT、MXM、FYM、FMS、FTK）の各人に関し、5240語のデータベースの偶数番単語から500語をランダム

に選択した（データベースについては、従来技術20参照。）。選択されたデータは、総計で3000語となり、話者依存型の各実験に使用したものとほぼ同数である。複数話者のHMMや話者依存型HMMの場合、明らかにより多量の学習データが必要であるが、当実験の目的は主として、デバッキングのためであった。HM網の生成処理は、話者依存実験の場合と同じ処理を使用した。すなわちHMMを、状態数200の場合は単一ガウス（混合数1）分布で、また状態数400では混合数1及び3で保持した。音声認識に際しては、話者6名の各人に付きランダムに選んだ100語を使用して、複数話者モードでテストした。

【0136】図9にその結果を示す。図9から明らかなように、本実施形態のSI－SSS法は、一貫して従来例のSSS法より良い結果を示している。その差が最も大きいのは、単一（混合数1）の混合ガウス分布を用いたより高いコンテキスト方向の分解度を有するHMM（状態数400）の場合である。混合数3を使用する場合はこの差が小さくなるが、このことは驚くことではない。なぜならば、混合数を複数にすることによって、異音の欠落を補償することができるためである。

【0137】

【発明の効果】以上詳述したように本発明に係る請求項1記載の不特定話者モデル生成装置によれば、複数の特定話者の発声音声データに基づいて話者独立型の隠れマルコフモデルを生成するモデル生成手段を備えた不特定話者モデル生成装置において、上記モデル生成手段は、複数の特定話者の発声音声データに基づいて、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布の隠れマルコフモデルを生成した後、上記単一ガウス分布の隠れマルコフモデルにおいて、1つの状態をコンテキスト方向又は時間方向に分割したときに、最大の尤度の増加量を有する状態を分割することを繰り返すことにより話者独立型の隠れマルコフモデルを生成する。従って、多数の話者の膨大な学習用テキストデータを必要とせず、従来例に比較して処理装置のメモリ容量の少なくてすみ、その計算時間を短縮することができる。

【0138】また、請求項2記載の不特定話者モデル生成装置においては、請求項1記載の不特定話者モデル生成装置において、上記モデル生成手段は、複数の特定話者の発声音声データに基づいて、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布の隠れマルコフモデルを生成する初期モデル生成手段と、上記初期モデル生成手段によって生成された単一ガウス分布の隠れマルコフモデルにおいて、1つの状態をコンテキスト方向又は時間方向に分割したときに、最大の尤度の増加量を有する状態を検索する検索手段と、上記検索手段によって検索された最大の尤度の増加量を有する状態を、最大の尤度の増加量に対応するコンテキスト方向又は時間方向に分割した後、バーム・ウェルチの学習アルゴリズムを用いて単一ガウス分布の隠れマルコフモデルを生成する生成手段と、上記生成手段の処理と上記検索手段の処理を、単一ガウス分布の隠れマルコフモデル内の状態を分割することができなくなるまで又は又は単一ガウス分布の隠れマルコフモデル内の状態数が予め決められた分割数となるまで繰り返すことにより、話者独立型の隠れマルコフモデルを生成する制御手段とを備える。従って、多数の話者の膨大な学習用テキストデータを必要とせず、従来例に比較して処理装置のメモリ容量の少なくてすみ、その計算時間を短縮することができる。

【0139】さらに、請求項3記載の不特定話者モデル生成装置においては、請求項2記載の不特定話者モデル生成装置において、上記検索手段によって検索される状態は、直前の処理で上記生成手段によって分割された新しい2つの状態に限定される。これによって、請求項1又は2記載の装置に比較して処理装置の計算時間を短縮することができる。

【0140】さらに、請求項4記載の不特定話者モデル生成装置においては、請求項2記載の不特定話者モデル生成装置において、上記検索手段によって検索される状態は、直前の処理で上記生成手段によって分割された新しい2つの状態と、上記新しい2つの状態から距離が1だけ離れた状態とに限定される。これによって、請求項1、2又は3記載の装置に比較して処理装置の計算時間を短縮することができる。

【0141】またさらに、請求項5記載の音声認識装置においては、入力される発声音声文の音声信号に基づいて所定の隠れマルコフモデルを参照して音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項1乃至4のうちの1つに記載の不特定話者モデル生成装置によって生成された話者独立型の隠れマルコフモデルを参照して音声認識する。従って、生成された不特定話者モデルを参照して音声認識することができ、従来例に比較して音声認識率を改善することができる音声認識装置を提供することができる。

【図面の簡単な説明】

【図1】　本発明に係る一実施形態である音声認識装置のブロック図である。

【図2】　図1の不特定話者モデル生成部31によって実行されるSI－SSS法の話者モデル生成処理を示すフローチャートである。

【図3】　図2のステップS15で用いる最尤分割設定処理のサブルーチンを示すフローチャートである。

【図4】　従来例及び実施形態における状態分割を示すHM網の一例の状態遷移図であって、（a）は元のHM網であり、（b）はコンテキスト方向の分割を示すHM網であり、（c）は時間方向の分割を示すHM網である。

【図5】　図2のSI－SSS法の話者モデル生成処理によって実行される時間方向の分割を示す状態遷移図で

ある

【図６】　図２のＳＩ－ＳＳＳ法の話者モデル生成処理によって実行される時間方向の分割のために、数５２及び数５３を用いてパラメータｆｈｔ（ｑ）及びｍｈｔ（ｑ，ｑ′）を計算するときに用いるデータ及び状態を示す図である。

【図７】　話者依存型実験のための初期ＨＭ網のトポロジーを示す状態遷移図である

【図８】　図２のＳＩ－ＳＳＳ法の話者モデル生成処理と図１３のＳＳＳ法の話者モデル生成処理に対するＣＰＵ時間の比較を示すグラフである。

【図９】　図１の音声認識装置を用いたときの、複数の話者認識タスクに対する図２のＳＩ－ＳＳＳ法の話者モデル生成処理と図１３のＳＳＳ法の話者モデル生成処理の音素認識率を示すグラフである。

【図１０】　従来例のＨＭ網の構造を示す状態遷移図である

【図１１】　従来例のＨＭ網で表現される各異音モデル構造を示す状態遷移図である

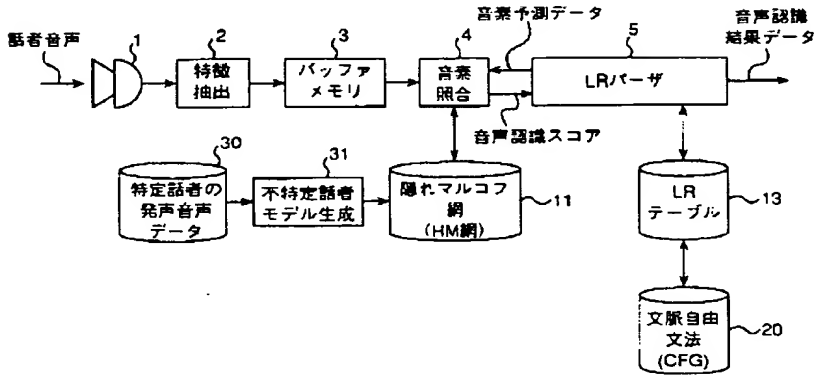
【図１２】　従来例のＳＳＳ法の話者モデル生成処理の原理を示す図である

【図１３】　従来例のＳＳＳ法の話者モデル生成処理を示すフローチャートである。

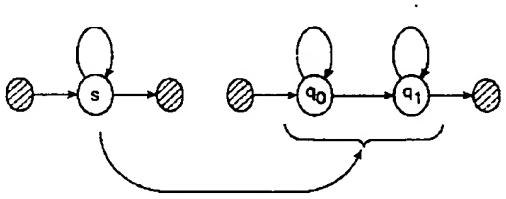
【符号の説明】

１…マイクロホン、
２…特徴抽出部、
３…バッファメモリ、
４…音素照合部、
５…ＬＲパーザ、
１１…隠れマルコフ網（ＨＭ網）、
１３…ＬＲテーブル、
２０…文脈自由文法データベース、
３０…特定話者の発声音声データ、
３１…不特定話者モデル作成部

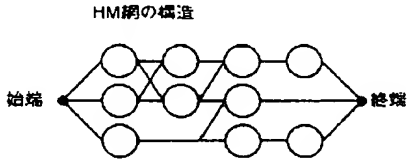
【図１】



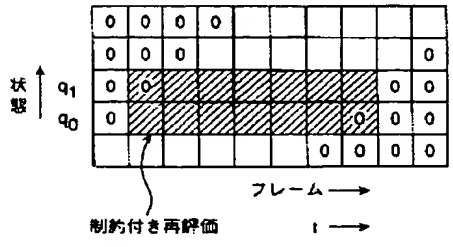
【図５】



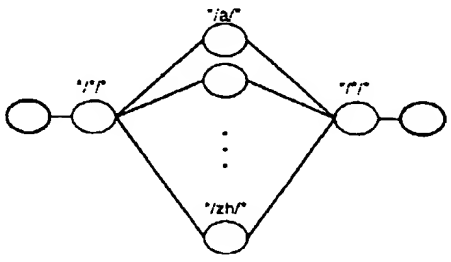
【図１０】



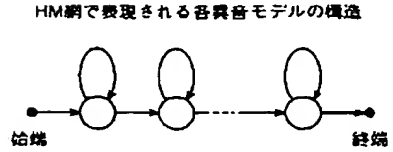
【図６】



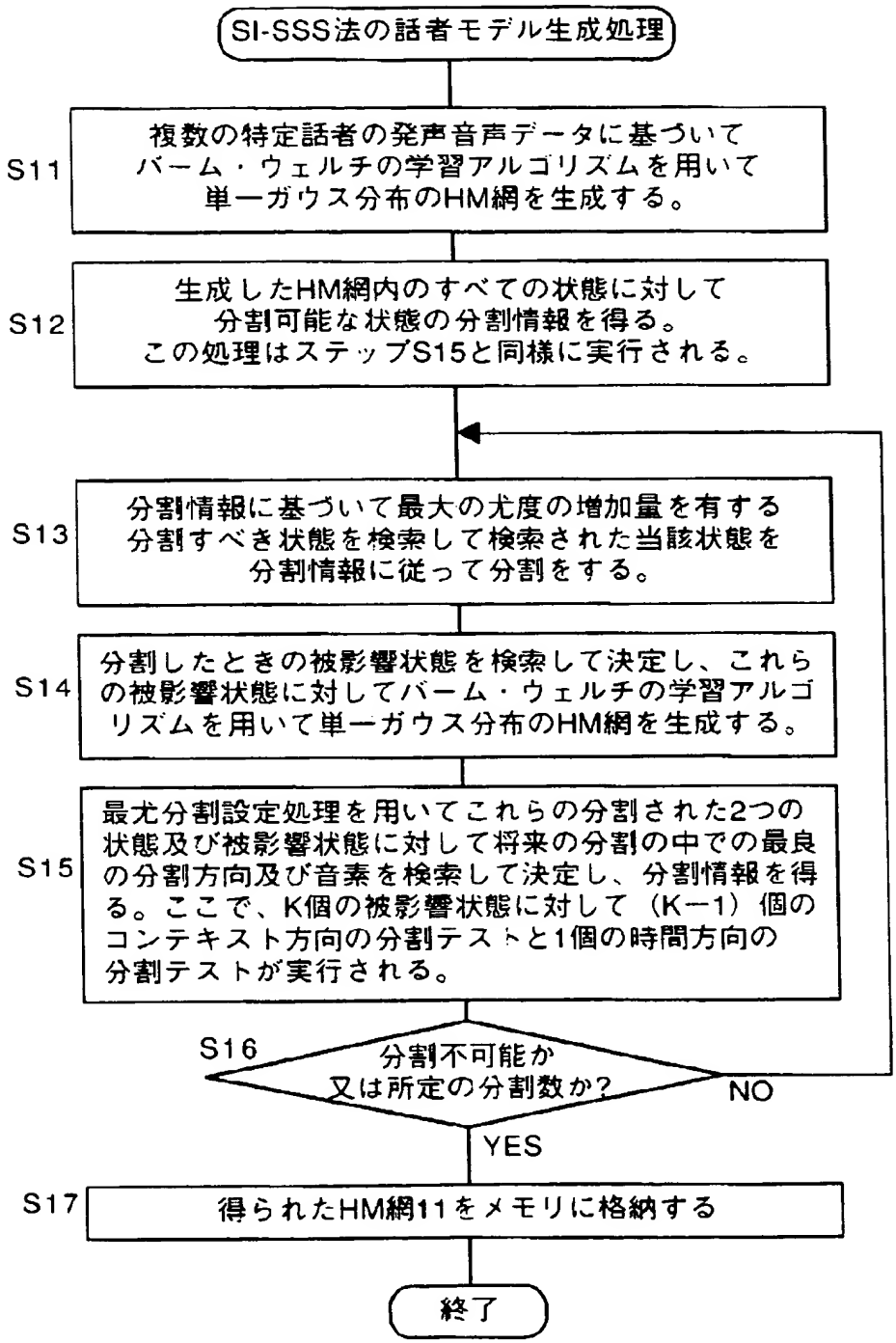
【図７】



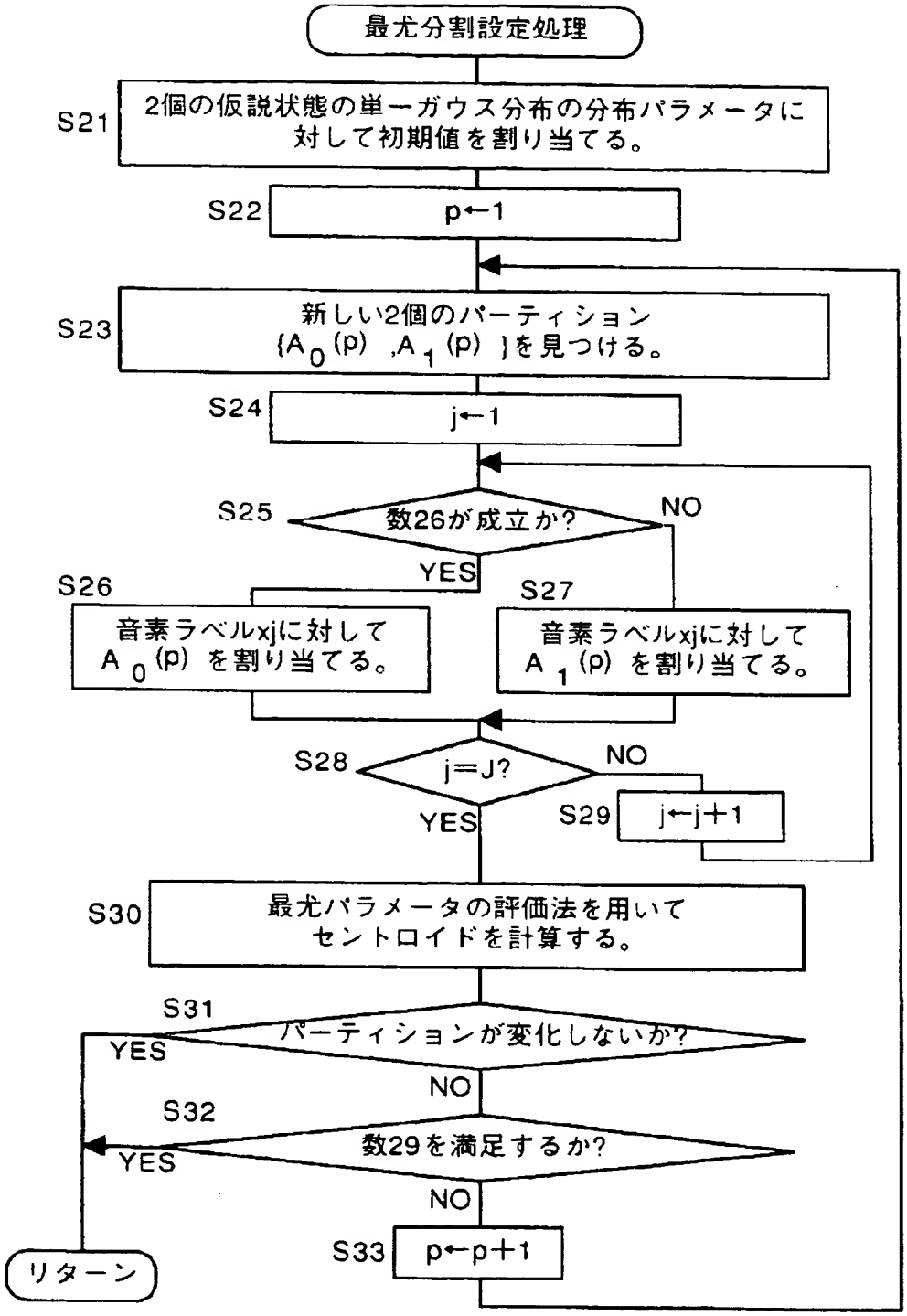
【図１１】



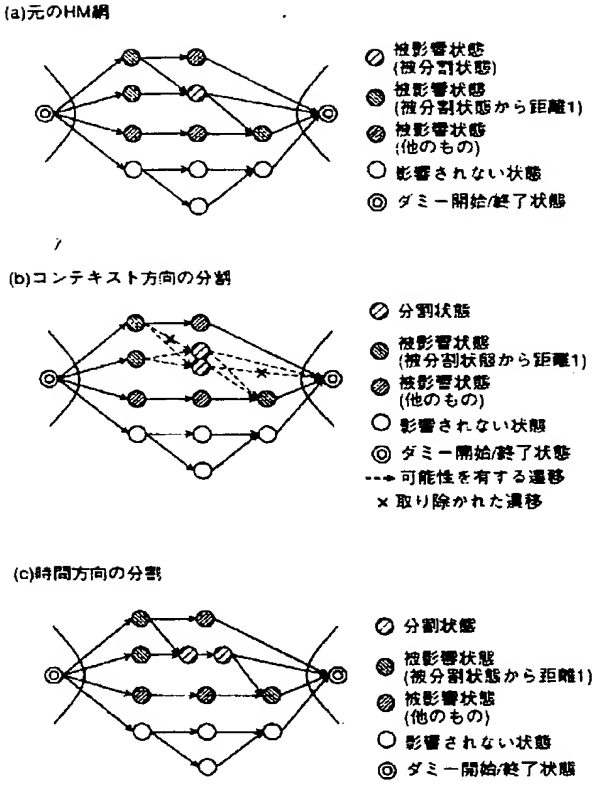
【図2】



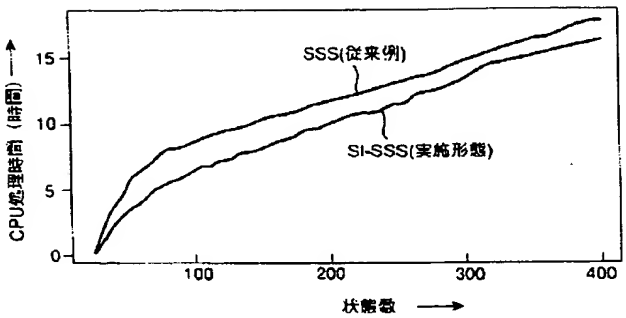
【図3】



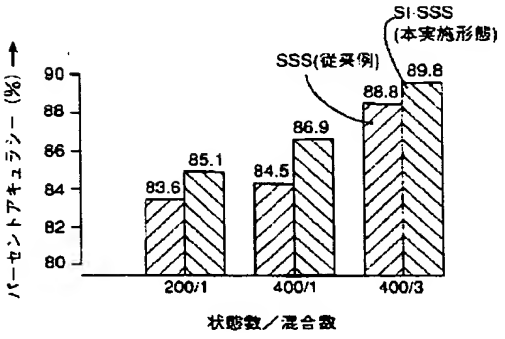
【図4】



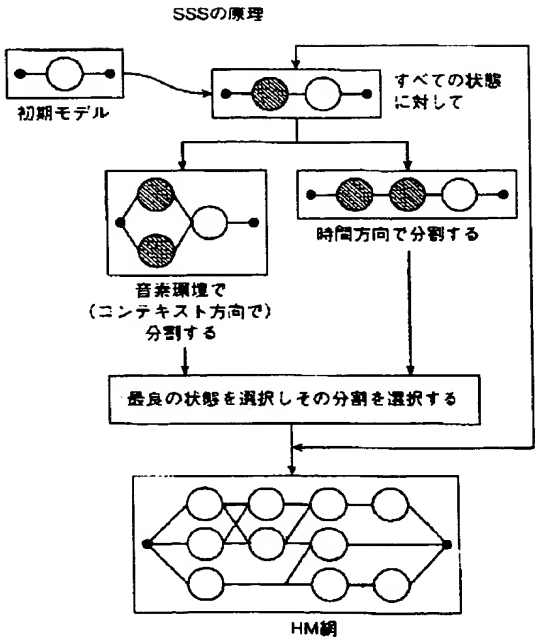
【図8】



【図9】



【図12】



【図13】